ARI Research Note 90-104

AD-A226 130

Optimizing the Long-Term Retention of Skills: Structural and Analytic Approaches to Skill Maintenance

Alice F. Healy, K. Anders Ericsson, and Lyle E. Bourne, Jr.

University of Colorado

for

Contracting Officer's Representative
Judith Orasanu

Basic Research
Michael Kaplan, Director

August 1990

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON Technical Director	JON W. BLADES COL, IN Commanding

---

Research accomplished under contract for the Department of the Army

University of Colorado

Technical review by

George W. Lawton

Accesion For	
NTIS CRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS —
2a. SECURITY CLASSIFICATION AUTHORITY — 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE —	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) 153-0638	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 90–104
6a. NAME OF PERFORMING ORGANIZATION University of Colorado 6b. OFFICE SYMBOL (If applicable) —	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code) Campus Box B-19 Boulder, CO 80309	7b. ADDRESS (City, State, and ZIP Code) Office of Basic Research 5001 Eisenhower Ave (PERI-BR) Alexandria, VA 22333-5600
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U.S. Army Research Institute 8b. OFFICE SYMBOL (If applicable) PERI-BR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903–86-K-0155
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333–5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102B; PROJECT NO. 74F; TASK NO. NA; WORK UNIT ACCESSION NO. NA

11. TITLE (Include Security Classification)
Optimizing the Long-Term Retention of Skills:
Structural and Analytic Approaches to Skill Maintenance

12. PERSONAL AUTHOR(S)
Healy, Alice F.; Ericsson, K. Anders; and Bourne, Lyle E., Jr.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 86/08 TO 90/02	1990, August	135

16. SUPPLEMENTARY NOTATION
Judith Orasanu, contracting officer's representative

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Retention; Skill building; Long-term memory; Memory; Electronics

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This research program seeks to identify the characteristics of knowledge and skill which are most resistant to decay due to disuse. Our research can be divided into two complementary parts. The first part is concerned with experimental analysis of factors influencing and improving retention of skill components. The second part is concerned with analysis and assessment of the structure of acquired memory and skills and how to monitor differential retention of components. For the analytic approach we developed five laboratory methodologies, and we completed investigations for each of them. We also identified four natural skills and completed investigations for each of them. For the structural approach we designed an experimental paradigm which allows us to assess the detailed encoding of new knowledge at presentation and at delay using verbal report techniques and chronometric measurement of retrieval components. Several studies of retention of vocabulary items were completed. In a number of our lines of investigation, we found evidence for a surprising degree of long-term skill retention. We formulated a theoretical framework, focusing on the importance of procedural reinstatement, and this framework enables us to understand this impressive memory performance. In contrast, in other studies we have

(Continued)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Judith Orasanu	22b. TELEPHONE (Include Area Code) (202) 274–5590 22c. OFFICE SYMBOL PERI-BR

DD FORM 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

ARI Research Note 90-104

19. Abstract (Continued)

conducted, we found considerable forgetting over even relatively short retention intervals. We have been able to place these studies in the same general theoretical framework developed to account for permastore, and we have been able to derive from these studies indications of the specific factors which facilitate retention. This document also includes two reports: (1) Optimizing the Long-Term Retention of Skills: Analytic and structural Approaches to Skill Maintenance: Review of Progress on Analytic Approach by Alice Healy and Lyle E. Bourne, Jr., and (2) The Long-Term Retention of Skills by Alice F. Healy, David W. Fendrich, Robert J. Crutcher, Wiliam T. Wittman, Antoinette T. Gesi, K. Anders Ericsson, and Lyle E. Bourne, Jr.

Optimizing the Long-term Retention of Skills:
Structural and Analytic Approaches to Skill Maintenance

Alice F. Healy, K. Anders Ericsson, and Lyle E. Bourne, Jr.

This research program seeks to identify the characteristics of knowledge and skill which are most resistant to decay due to disuse. Our research can be divided into two complementary parts. The first part is concerned with experimental analysis of factors influencing and improving retention of skill components. The second part is concerned with analysis and assessment of the structure of acquired memory and skills and how to monitor differential retention of components. The eventual goal of both parts is to be able to make relevant recommendations about training routines for long-term skill maintenance.

A new line of investigation, involving both the analytic and structural approaches, began consequent to the arrival of three Army tank simulators. This effort is concerned with the study of complex military skills. A study was completed involving the extensive training of two subjects with the simulators.

The analytic approach. We have developed two lines of research for investigating skill retention and maintenance using the analytic approach. The first line of research involves investigating different laboratory analogues of component skills of electronic technicians. The second complementary line of research involves investigating parallel natural skills learned by the college population during their prior education.

We have developed five laboratory methodologies, and we have completed investigations for each of them. The laboratory tasks involve (a) target detection, (b) data entry, (c) learning logical rules involved in circuit design, (d) memory for numerical calculations, and (e) temporal, spatial, and item components of memory for lists. We have also identified the following four natural skills and have completed investigations for each of them: (a) mental multiplication, (b) algebra, (c) data entry, and (d) temporal, spatial, and item components of memory for class schedules.

The structural approach. We have designed an experimental paradigm which allows us to assess the detailed encoding of new knowledge at presentation and at delay using verbal report techniques and chronometric measurement of retrieval components. Several studies of retention of vocabulary items have been completed, in which subjects have been instructed to use the keyword method with supplied keywords.

Overview of findings. One of our initial aims in this research was to validate the concept of "permastore" initially proposed by Bahrick to account for his finding that some information remains permanently intact in memory. Indeed, in a number of our lines of investigation (e.g., target detection, mental multiplication, and data entry), we have found evidence for a surprising degree of long-term skill retention. We have formulated a theoretical framework, focusing on the importance of procedural reinstatement, and this framework enables us to understand this impressive memory performance. In contrast, in other studies we have conducted (e.g., memory for numerical calculations, vocabulary learning, and components of memory for class schedules), we found considerable forgetting over even relatively short retention intervals. We have been able to place these studies in the same

general theoretical framework developed to account for permastore, and we have been able to derive from these studies indications of the specific factors which facilitate retention.

Optimizing the Long-term Retention of Skills:
Structural and Analytic Approaches to Skill Maintenance

A. Accomplishments

Proposal. At the beginning of this year, we prepared our contract renewal proposal, which we submitted on October 28, 1988. This proposal includes 29 experimental studies, 9 of which are specified in detail. Nine of the 29 experiments mentioned and 5 of the 9 experiments described in detail involve the TopGun simulator. The development of these experiments required that we master the programming and operation of this simulator. We also completed a preliminary experiment with the simulator. This experiment involved extensive training of two subjects for twelve acquisition sessions. A retention test followed three months after the end of acquisition.

We recently submitted an Addendum to our contract renewal proposal. In this Addendum we relate our proposal to the theoretical framework we subsequently developed and discussed at the ARI contractors' meeting in Ft. Gordon, Georgia (see below). We also proposed two new experiments that would provide direct tests of our theoretical framework and extend its generalizability.

Meetings and visits. In September, Dr. Walter Schneider, a fellow ARI contractor from the University of Pittsburgh, visited our campus and our laboratory. Both as a group and individually the members of our project discussed with him our research efforts and learned about his related efforts. Dr. Schneider provided some very helpful and stimulating feedback to us.

In September, Bourne visited the University of Wisconsin and discussed our research on data entry at a symposium there.

In November, eight of us (Healy, Bourne, Ericsson, Fendrich, Crutcher, Messamer, Frick, and Tetewsky) went to the annual meeting of the Psychonomic Society in Chicago, Illinois. We discussed our research with colleagues both informally and in two formal presentations.

Three of us (Healy, Ericsson, and Bourne) attended the annual contractors' meeting on Skill Acquisition and Retention at Ft. Gordon, Georgia on February 27 - March 1. We presented a report at that meeting which summarized some of our recent experimental findings and placed them into a general theoretical framework which allowed us to account for both the impressive memory performance we have observed in the target detection, mental multiplication, and data entry tasks as well as the considerable forgetting we have observed in our studies of memory for numerical calculations, vocabulary learning, and components of memory for course schedules. We also included in the presentation a discussion of our initial work with the TopGun simulator and some of the research we have planned with the simulator.

At the meeting in Georgia, we met Dr. Robert Wisher, an ARI researcher who introduced us to an intriguing problem of individual differences involving the training of receivers of Morse Code. We discussed this problem with him, he subsequently sent us some relevant material, and we initiated plans to investigate possible ways of solving this problem both in our laboratory and in collaboration with Dr. Wisher at the training site itself. One of the two new

studies in our addendum involves the task of Morse Code. This study investigates the long-term retention of that task but may have interesting implications concerning the problem of individual differences.

In August, two of us (Healy and Bourne) attended the annual meeting of the American Psychological Association. Healy presented an invited address entitled "The long-term retention of skills." This address summarized the work on our project. A copy of the paper is attached here (see Appendix A).

Technical reports. At the beginning of this year, we performed the final revisions on two technical reports summarizing our work in this project during its initial stages when it was supported by the Air Force Human Resources Laboratory, before support was provided by the ARI. A copy of each of these reports is attached here (see Appendix B and Appendix C).

TopGun simulators. We completed two sets of analyses of our preliminary experiment with the TopGun Simulators. This experiment involved extensive training of two subjects followed by a long-term retention test one-month at the end of acquisition. The initial analyses employed a number of different indices of performance. According to all of the indices, there was substantial improvement in the skill across the acquisition sessions. According to some but not all of these indices, there was significant forgetting across the retention interval, but even in those cases performance at the retention test was superior to that exhibited at the beginnning of acquisition.

The aim of the second set of analyses was to isolate the various processing components of the tank gunnery skill. Indeed, we were able to distinguish between some components which showed forgetting across the three-month retention interval and other components which showed no forgetting. However, these analyses can only be seen as suggestive, because of the small number of subjects employed so far in this experiment.

The Analytic Approach

We have made further progress in our testing of both the laboratory skills and the natural skills which we began in the first two years.

Laboratory Skills

Target detection. After a delay of approximately six months, we retested our third and fourth subjects given extensive training in target detection. These subjects were trained by means of a varied mapping procedure, instead of the consistent mapping procedure used with our first two subjects. One of these new subjects showed very little learning during training, so that his retention data are of minimal interest. The other subject, however, did show considerable improvements during acquisition, so that it is of interest to determine how much forgetting she demonstrated over the long retention interval. We found very little forgetting, an amount roughly comparable to that of the first two subjects. Because this new subject was exposed to varied mapping training whereas our first two subjects were exposed to consistent mapping training, two interesting tentative conclusions can be drawn. First, consistency of training does not seem to be necessary for entry into permastore. Second, automaticity does not seem to be associated with entry into permastore. This latter conclusion follows from the fact that previous researchers, including Shiffrin and Schneider (1977), found no evidence of automaticity with varied mapping

training.

We also completed for a fifth subject the initial acquisition phase of extensive training on the target detection skill with the varied mapping procedure. Preliminary data analysis revealed that this subject also showed improved performance both in terms of response latencies and accuracy across the twelve sessions of training. Further, a decrease in the frame size effect indicated that this subject became more automatic as training progressed. This subject will be retested next year in order to determine how well this target-detection skill is retained.

In collaboration with Janet Proctor, we conducted a new experiment which followed our observation that the word frequency disadvantage (the tendency to miss target letters in common as opposed to rare words) was diminished with previous practice at detecting letters in prose. For this new experiment we constructed a prose passage which we could use in three different letter detection tasks, one involving the letter n and the common word and, the second involving the letter t and the common word the, and the third involving the letter h and the common word the. Subjects read this passage followed by a test passage involving the letter h and the common word the. We found that only practice with the same target letter yielded the reduction of the word frequency disadvantage. Practice with the same passage, even with the same common test word, was not sufficient if the subject did not have the same target letter. These findings were both clear-cut and surprising.

With Janet Proctor we later completed a second follow-up experiment investigating the loss of the word frequency disadvantage after practice performing a prose letter detection task. This experiment included important controls missing from the previous studies. The preliminary analyses of this experiment have recently been completed. They indicate that there is transfer from one prose letter-detection task to a second task involving a different target letter if and only if that letter also occurs in the same frequent test word and is processed more rapidly than the initial target letter. In particular, practice detecting the letter h, which occurs in the medial position of the word the, transfers to detecting the letter t, which occurs in the initial position of the word the, but does not transfer to detecting the letter n, which occurs in the medial position of the word and. Also, practice detecting the letters n or t does not transfer to detecting the letter h. These findings can be understood by considering that the letter t is probably detected before the letter h in the word the, and these results are consistent with our previous experiments, which had also indicated differences between the detection of the letters t and h.

On the basis of encouragement and suggestions by the journal editor and reviewers, we (Healy, Fendrich, and Proctor) completed three revisions of our manuscript reporting our initial studies of target detection; this manuscript is now in press in the Journal of Experimental Psychology: Learning, Memory, and Cognition. A copy of the manuscript is attached here (see Appendix D.) Our revision includes our new theoretical analysis of the factors influencing the degree of long-term retention. This analysis emphasizes the importance of procedural memory in enhancing skill retention.

Data entry. We prepared a talk summarizing three of our studies in our series of experiments on data entry. An early version of the talk was presented in September by Bourne at the University of Wisconsin. A subsequent version of

the talk (by Fendrich, Healy, and Bourne) was presented at the annual November meeting of the Psychonomic Society in Chicago. A copy of the Psychonomics talk is attached here (see Appendix E).

David Fendrich completed his doctoral dissertation, which is a report of three of our exprepriments on data entry. The dissertation also includes an extensive review of the literature on implicit memory. This dissertation was extremely well received by the examining committee, David Fendrich successfully passed his final oral examination, and his final version of the dissertation was approved by his committee and accepted by the Graduate School. A copy of the dissertation is attached here (see Appendix F). Dr. David Fendrich began work as a Postdoctoral Fellow at New York University in August.

We also completed the analyses and initial write-up of a new data entry experiment. This study provided the basis for an undergraduate honors thesis by Antoinette Gesi, who graduated this winter Magna Cum Laude. We (Gesi, Fendrich, Healy, and Bourne) also presented a paper describing this work at the joint annual meeting of the Rocky Mountain Psychological Association and the Western Psychological Association on April 27, 1989. A copy of this paper is attached here (see Appendix G). Antoinette Gesi will begin graduate work at the University of California, Santa Cruz, in the Fall.

Memory for numerical calculations. Our manuscript on this topic was accepted for publication with revisions suggested. We completed the recommended revisions and proofreading and the article is now in print in the Journal of Experimental Psychology: Learning, Memory, and Cognition. A copy of this article is attached here (see Appendix H).

Learning logical rules involved in circuit design. We completed the initial analyses of our follow-up study on this topic. Although the preliminary results are very promising, the initial analysis revealed that some subjects needed to be replaced because they did not successfully learn the rules in the time allotted. We finished testing the additional subjects needed but we have not yet completed the data analyses.

Temporal, spatial, and item components of memory for lists. In our contract renewal proposal we outlined a new series of experiments on this topic. We completed conducting two preliminary experiments in this series. These experiments compare memory for size order with memory for temporal and spatial order information. This work provided the basis for the first-year research project of USAF Captain Michael Scheall, who presented a summary of this study at the Department of Psychology Annual Miniconvention on May 1, 1989.

We also initiated the design of two new experiments that link our previous work on the learning and long-term retention of temporal, spatial, and item information with our previous work on data entry. These experiments should provide information concerning both the learning of a data entry skill when based on processing either temporal, spatial, or item information and the relative importance of the cognitive and the motor aspects of the long-term retention of data entry skills with respect to temporal, spatial, and item information.

Natural Skills

Mental multiplication. We retested both of the subjects given long-term

training in the multiplication skill, the first after a 7-month and then a 14-month retention interval (she had been retested initially after a 3-month interval) and the second after a 4-month and then a 7-month interval. At these tests both subjects showed very little loss of either speed or accuracy on this task, despite the fact that they had shown substantial gains in performance during the 12 acquisition sessions. Although retention was high in each case, some forgetting was evident. Recently, the first subject was retested again after an interval of approximately 20 months. At this retesting the original motor response was employed; at all of the previous retests an oral response was used instead. Forgetting was evident at this final retesting, but performance was still considerably better than at the start of training. Specifically, the response latencies at this retesting were comparable to those during the sixth training session.

We completed conducting a group experiment to assess the long-term retention of the improvements in multiplication performance learned in the laboratory. This experiment allowed us to assess whether subjects represent the two versions of a problem (e.g., 2 X 3 and 3 X 2) separately in memory or as a single entry. Our intial analyses revealed that the two versions of a problem are indeed stored separately, because subjects responded more rapidly and more accurately to the versions that they studied ("old" problems) during the acquisition session one-month previously than to the versions seen only at test ("new" problems). The one exception to this finding occurred with the problems involving the digit 1 (e.g., 1 X 3), which according to previous research do not rely on memory storage but rather the use of a simple rule. The finding for the problems involving 1 seems to rule out enhanced perceptual processing of the stimulus as the locus of the old-new difference. Rather, it seems that the association between the stimulus problem and the answer has been strengthened in memory. In any event, subjects showed no explicit recognition memory for the old problems, although the recognition memory test was given at the conclusion of the experiment, after the testing (which included both old and new problems) was completed.

We recently completed a second group experiment assessing the long-term retention of the improvements in multiplication performance learned in the laboratory. In this new study, subjects were trained on a subset of the problems comprising the 1-9 multiplication table. This experiment allowed us to assess the degree of transfer between related multiplication problems. For example, subjects were trained on 4 X 3 but not 3 X 4, 6 X 2, or 2 X 6; in a final retention test performance on all four problems was compared. A formal recognition procedure was also included in this experiment for some subjects at the beginning of the retention test and for other subjects at the end of the retention test. The initial analyses of this experiment revealed retention of the old problem reflected both in terms of improved response latencies relative to new problems and in terms of explicit recognition ratings. There was also significant transfer in terms of both of these measures to the reverse problems, which were the same as the old problems except for the order of the operands. This finding held even for those sets of problems which shared the same answer, as in the example above, so that the advantage for the reverse problems relative to other new problems could not be attributed simply to the fact that the reverse and old problems shared the same response. In agreement with the earlier study, subjects were not able to discriminate the old from the reverse problems on the explicit recognition test given at the end of the retention session, although such discrimination was evident on the recognition test given at the beginning of the retention session.

We (Fendrich, Healy, and Bourne) have submitted an abstract for a poster summarizing the two group experiments on mental multiplication for presentation at the next annual meeting of the Psychonomic Society.

Algebra skills. We (Meiskey, Healy, and Bourne) completed our invited report for the University of Colorado journal On Teaching. This article is now in press and scheduled to appear in print this Fall. A copy of this manuscript is attached here (see Appendix I).

Temporal, spatial, and item components of memory for course schedules. We completed the analyses of our second experiment on memory for course schedules, and we completed the second retesting of subjects in our first experiment. This study of memory for course schedules provided the basis for the doctoral dissertation of USAF Major William Wittman. The dissertation was completed and accepted with much praise by the committee and the graduate school this June. A copy of this dissertation is attached here (see Appendix J). Dr. Wittman joined the faculty of the U. S. Air Force Academy this summer. The following is a brief summary of the findings and conclusions from this study:

Results of our first experiment demonstrated the superior retention of the spatial component over the temporal or item components in memory for naturally learned course schedules. Spatial component recall showed a clear advantage for all three test sessions, with retention intervals ranging from 12 to 36 months. This long-term advantage for spatial information is consistent with the short-term advantage for spatial information we have found the laboratory.

Our second experiment involving memory for course schedules confirmed this spatial component superiority. In this experiment, we sought to determine what degree proceeding to and from classes contributed to the spatial advantage. Subjects were therefore instructed to learn an actual class schedule, but were never given the opportunity to physically practice their schedule, that is, to walk from building to building. Results again showed a spatial component advantage. Additionally subjects were asked both to locate their classes on a campus map and to provide the names of class locations. In the naming of class locations, subjects did not show an advantage for this information over the temporal or item components, thus suggesting a truly spatial memory interpretation of results.

Both experiments provide support for the general theoretical framework we have applied in our other experiments. That is, memory performance can be explained in terms of the degree to which the demands of the memory test allow reinstatement of procedures acquired in training. Such procedures may be both motoric (as in walking to a class) and cognitive (as acquired in studying the campus map).

Data entry. We retested our single subject who had extensive practice with the data entry skill. This retesting took place after a 21-month retention interval. Although significant forgetting was found, most impressive was the extremely high degree of performance which was maintained across the lengthy retention interval.

The Structural Approach

We conducted a number of studies to extend our previous investigations of the acquisition and retention of Spanish-English vocabulary items.

Specifically, we completed a study to assess the effect of extended practice on retrieval of the English equivalents of Spanish words. Our primary goal was to determine whether a direct association is formed between the Spanish word and the English equivalent as a result of extended practice. The initial results suggest that subjects are indeed able to retrieve directly the English equivalent from the Spanish word.

We also designed the materials for a new experiment to examine how preexisting knowledge can be used initially to acquire and later to reaccess Spanish vocabulary items.

In addition, we recalled four subjects from our original vocabulary retention study and retested them after a one-year delay. Retention of the English equivalents declined considerably: from 80% (one week or one month delay) to less than 15% (after one year). However, retention results for the two subtasks, the Keyword Retrieval Task (Spanish-Keyword) and the English Retrieval Task (Keyword-English), were strikingly disparate. Performance on the English Retrieval Task was similar to the Full Retrieval Task: less than 15% recall. However, performance on the Keyword Retrieval Task was 80% correct, almost no loss at all. These results, consistent with our earlier ones, suggest that failure to recall the English equivalents is due to difficulty in recalling the interactive image. The keyword component, on the other hand, shows remarkably good retention, even after a delay of a year.

Finally, we completed a keyword generation experiment, in which we asked subjects to generate keywords for the Spanish words used in the retention experiments. Between 40% and 60% of the time, subjects in the generation experiment generated the same keywords as we had used in the retention experiment. Comparing these numbers to the 80% recall after a year suggests that the remarkable retention for keywords is in part due to inference processes. Subjects either guess the keyword in some cases or are able to generate likely keyword candidates and then check the memory trace. Further, we completed running a second group of subjects in the keyword generation experiment. These data have been tabulated but have not yet been analyzed.

Optimizing the Long-term Retention of Skills: Structural and Analytic Approaches to Skill Maintenance

A. Accomplishments

Meetings. On September 8, Healy delivered the first colloquium of the year at the Institute of Cognitive Science, University of Colorado, Boulder. In that talk she reviewed the progress made on this project concerning the long-term retention of skills.

On September 14, Healy delivered a colloquium at Wellesley College. She was the first speaker of the year in their Cognitive Science Colloquium Series. In that talk she also reviewed our research on the long-term retention of skills.

On September 16, Healy delivered an invited talk at Harvard University at the Symposium in Honor of William K. Estes. Again Healy spoke about the long-term retention of skills.

On October 11, Bourne delivered a Colloquium at Loyola University. He discussed our research in this project with the data entry task.

The Analytic Approach

We have made further progress in our testing of both the laboratory skills and the natural skills which we began in the first twelve quarters. We also designed and completed the computer programming for an experiment examining the acquisition and long-term retention of the skill of receiving Morse code. This experiment was outlined in the addendum to our ARI contract renewal proposal which we submitted this summer. This experiment is scheduled to be conducted during the next quarter. The experiment is being performed by Deborah Clawson, a new graduate student who joined our laboratory this fall.

Laboratory Skills

Target detection. After a delay of approximately four months, we retested our fifth subject given extensive training on the target detection skill with the varied mapping procedure. The data from this retention test have not yet been analyzed.

We completed a set of analyses on our second follow-up experiment investigating the loss of the word frequency disadvantage after practice performing a prose letter detection task. This experiment was conducted in collaboration with Janet Proctor at Purdue University. These analyses confirm the preliminary analyses reported earlier and indicate that the loss of the word frequency disadvantage cannot be attributed to the subjects' adopting a strategy in which they pay special attention to common function words like the and and.

Also in collaboration with Janet Proctor we designed and conducted a third follow-up experiment investigating the loss of the word frequency disadvantage. This experiment was also conducted at Purdue University. In this experiment we compared three different practice situations. All three situations involved the letter-detection task; they differed in the type of passage subjects read. The passage either consisted of meaningful prose (as in the previous studies),

scrambled words which did not form a meaningful text, or scrambled letters which did not include real words. This experiment should enable us to assess whether practice with text comprehension is necessary for loss of the word frequency disadvantage. We will begin tabulating and analyzing the data from this experiment during the next quarter.

Further in collaboration with Janet Proctor we designed a fourth follow-up experiment investigating the loss of the word frequency disadvantage. In this experiment subjects will perform the letter detection task either on a computer terminal or with paper and pencil. Both of these procedures have been employed in the past, but we have never previously examined the effects of transferring from one procedure to the other, which will be the focus of this experiment. We are currently doing the computer programming necessary to conduct this experiment.

Data entry. In collaboration with Dr. David Fendrich, who is now a Postdoctoral Fellow at New York University, we began preparing a manuscript for publication in a major journal which summarizes three of our experiments with the data entry task. These experiments formed the basis of Dr. Fendrich's dissertation.

In colloaboration with Antoinette Gesi, who is now a graduate student at the University of California, Santa Cruz, we began preparing a manuscript for publication in a major journal which summarizes another one of our experiments with the data entry task. This experiment formed the basis of Ms. Gesi's undergraduate honors thesis.

We completed a pilot study examining the effects of increasing the difficulty of the data entry task by requiring subjects to perform a concurrent task involving articulatory suppression. The effects of transferring from the easy version of the task (no suppression) to the hard version (with suppression) and transferring from the hard version to the easy version were also compared. This pilot study involved only two subjects, one of whom was given extensive training with the hard task and one with the easy task. The results were inconclusive because of the small number of subjects employed but showed interesting trends suggesting that practice in the difficult version leads to superior retention of the skill after a one-week delay. This study was conducted as part of a course research requirement by Vivian Schneider, who is an advanced graduate student in our laboratory.

Memory for numerical calculations. We designed a new experiment which combines the techniques we used in our study of memory for numerical calculations with those used in our study of mental multiplication. The computer programming necessary to conduct this experiment has been completed, and the experiment is scheduled to be conducted during the next quarter. The general outline of this study was included in the addendum to our ARI contract renewal proposal, which we submitted this summer. The experiment is being performed by Danielle McNamara, a new graduate student who joined our laboratory this fall.

Learning logical rules involved in circuit design. We have been retabulating our data from our most recent experiment on this topic so that we can conduct an analysis which is sensitive to the stimulus shown to each subject on each trial.

Temporal, spatial, and item components of memory for lists. We have been designing two new experiments in our series comparing memory for size order with memory for temporal and spatial order information. In these experiments we plan to manipulate the subjects' task both during stimulus presentation and during the retention interval. These manipulations are aimed to allow us to determine the coding strategies used by subjects in order to retain each of the three types of information. The new experiments will form the basis of a masters thesis by USAF Captain Michael Scheall.

Natural Skills

Mental multiplication. We have been preparing a report of our two most recent experiments on mental multiplication. This report will be presented as a poster during the next quarter at the annual meeting of the Psychonomic Society.

We designed and completed the computer programming for a new experiment on mental multiplication with the aim to evaluate the nature of the information learned by subjects who are given training in mental multiplication. Subjects will be given training with two different types of multiplication problems (e.g., __ = 3 x 6 and 28 = __ x 7), and we will examine the amount of transfer from one type of problem to the other. This experiment is scheduled to be conducted during the next quarter. It forms the basis of the first year research project of a new graduate student, Tim Rickard, who joined our laboratory this fall.

Algebra skills. We have no new progress to report on this topic.

Temporal, spatial, and item components of memory for course schedules. In collaboration with Major William Wittman, who is now on the faculty of the U.S. Air Force Academy, we have been preparing a manuscript for publication in a major journal which summarizes our work on components of memory for course schedules. This study formed the basis of the doctoral dissertation by Major Wittman.

Data entry. We have no new progress to report on this topic.

The Structural Approach

The research activity of the structural approach has been focused on two projects: retention of foreign vocabulary items and retention of acquired memory skills.

Retention of foreign vocabulary items. During the past quarter, Robert Crutcher has completed his masters thesis, which reports the complete results for two experiments on the encoding and retention of Spanish-English vocabulary items. The defense of Mr. Crutcher's masters thesis is scheduled for November 13, 1989.

Acquisition and retention of memory skill for dinner orders. During the late spring and summer of 1989 Deborah Bauder in collaboration with Anders Ericsson trained a subject (JH) to use the mnemonic encodings and retrieval structure of an expert waiter (JC), whose exceptional memory for dinner orders had been previously examined by Ericsson and Polson. After over 80 hours of practice the subject's ability to memorize dinner orders from tables of 3, 5, and 8 people had reached a level comparable to that of the expert waiter (JC).

At that point in time Anders Ericsson and Robert Crutcher conducted a series of three experiments, each consisting of six hour-long test sessions. The first experiment was a replication of the first experiment with the expert waiter (JC). The last two experiments examined JH's memory performance with different materials to examine under which conditions the newly acquired memory skill would not be transferable and his memory performance would return to the level of his original memory performance. The data from these experiments have been prepared for data entry and the statistical analyses are forthcoming. In a year we plan to bring back the subject for detailed examination of his retention of the acquired memory skill.

B. Problems Met or Anticipated

Our ARI contract has not yet been refunded. We hope that we will receive new funds in the 1990 fiscal year. We received a six-month no-cost extension of our contract, but at this point we have essentially exhausted all of our contract funds. In addition, three of our most energetic researchers, Dr. David Fendrich, Dr. William Wittman, and Ms. Antoinette Gesi, have left our laboratory because they received their academic degrees and we were unable to offer them employment here. Nevertheless, our laboratory has continued to be active. In large part this activity is due to the fact that we were able to attract three new graduate students and two new undergraduates to work with us with essentially no monetary compensation (the only one receiving monetary support from us this fall is one of the undergraduates who is paid on a work-study scholarship). Also, the University of Colorado has generously provided us with small grants that enable us to buy supplies and fix our equipment. Because of the uncertainty concerning our future ARI funding, we have decided to postpone our experiments with the TopGun simulators. We will initiate these crucial experiments as soon as we hear that our ARI funding will in fact be renewed.

C. Activities Anticipated During the Next Reporting Period

The analytic approach. We plan on continuing each of the studies of laboratory and natural skills discussed above.

The structural approach. We plan to continue the research on the projects mentioned above. In particular, we will initiate pilot studies of the effectiveness of retention for vocabulary items encoded with existing semantic knowledge.

D. Funds Expended and Remaining

The information about funds expended and remaining is provided in a separate report from the University of Colorado Office of Contracts and Grants.

E. Utilization of Personnel

Paid Graduate Student Research Assistants:

Robert Crutcher (20 hours per week summer only)

Undergraduate Work-Study Assistants:

Antoinette Gesi (40 hours per week professional research assistant summer only), Michael Kos (10 hours per week work-study), Bill Marmie (20 hours per week

work-study summer; 10 hours per week work study fall)

Volunteer Graduate, Undergraduate, and Postdoctoral Students:

Robert Crutcher (graduate student), Grant Sinclair (graduate student), Michael Scheall (graduate student), Lori Meiskey (graduate student), Vivian Schneider (graduate student), Danielle McNamara (graduate student), Deborah Clawson (graduate student), Tim Rickard (graduate student), Mark Gehman (undergraduate student), Sheldon Tetewsky (postdoctoral fellow).

F. Honors, Publications, and Conference Papers

Honors:

Elected Member-at-Large of the Executive Committee of Division 3, American Psychological Association (Healy); $7,100 Psychology Department Overhead Fund Grant (Healy, Ericsson, and Bourne)

Publications:

Crutcher, R.J., & Healy, A.F. Cognitive operations and the generation effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1989, 15, 669-675.

Ericsson, K.A. (1990). Theoretical issues in the study of exceptional performance. In K.J. Gilhooly, M. Keane, R.H. Logie, & G. Erdos (Eds.), Lines of thought (Vol. 2, pp. 5-28). London: Wiley.

Ericsson, K.A., & Crutcher, R.J. (1989). The nature of exceptional performance. In P.B. Baltes, D.L. Featherman, & R.M. Lerner (Eds.), Life-span development and behavior (Vol. 10, pp. 187-217). Hillsdale, N.J.: Erlbaum.

Schneider, V.I., Healy, A.F., Ericsson, K.A., & Bourne, L.E. Letter detection errors in reading, auditory, and memory tasks. Journal of Memory and Language, 1989, 28, 400-411.

Conference papers and colloquia:

Bourne, L.E. Long-term retention of procedural and episodic memory for digits. Colloquium delivered at Loyola University, Chicago, Illinois, October 11, 1989.

Healy, A.F. The long-term retention of skills. Colloquium delivered at the Institute of Cognitive Science, University of Colorado, Boulder, Colorado, September 8, 1989.

Healy, A.F. The long-term retention of skills. Invited address at the 97th Annual Meeting of the American Psychological Association. New Orleans, Louisiana, August 12, 1989.

Healy, A.F. How much do we remember of what we've learned?: The long-term retention of skills. Colloquium delivered at Wellesley College, Wellesley, Massachusetts, September 14, 1989.

Healy, A.F. The long-term retention of skills. Invited paper presented at

the Symposium in Honor of William K. Estes. Harvard University, Cambridge, Massachusetts, September 16, 1989.

Krampe, R.Th., Tesch-Romer, C., & Ericsson, K.A. Voraussetzungen fuer die Entwicklung von Hoechstleistungen [Prerequisites for the acquisition of expert-level performance]. 9th Tagung Entwicklungspsychologie, Munich, West Germany, September 18-21, 1989.

Optimizing the Long-Term Retention of Skills:

Analytic and Structural Approaches to Skill Maintenance

Principal Investigators: Healy, Ericsson, and Bourne

Review of Progress on Analytic Approach (Healy and Bourne)

Our project has two major lines of research, which have the same general aims but differ in their methodological approach. One line involves what we call the "analytic approach," which makes use of the classical tools of experimental psychology to investigate complex tasks by analyzing them into simpler component subtasks and then intensively studying the simpler components. The second line involves what we call the "structural approach", which makes use of verbal protocols and chronological measurement techniques to describe the structure of complex tasks. Lyle Bourne and I have been collaborating with our students using the analytic approach, whereas Anders Ericsson and his students have been using primarily the structural approach. Today I will first review the progress we have made in our studies with the analytic approach, and then Anders will discuss our progress with the structural approach.

Lest you think that these lines of investigation function independently, let me begin by saying that both have had as their overarching goal finding ways to optimize long-term retention, particularly the long-term retention of skilled performance. Our separate efforts beneficially inform each other. We all start with the assumption that some part of acquired knowledge or skill is permanent. Harry Bahrick demonstrated that, while a large part of acquired knowledge is lost rapidly, a significant proportion can last a lifetime, even if that knowledge is not intentionally rehearsed or accessed in the meantime. We adopt Bahrick's concept of a "permastore" as a fundamental fact of memory, and we look for conditions of training or attributes of learned material that lead to permastore. Indeed, in a number of our studies — those involving the tasks of

target detection, mental multiplication, and data entry -- we have found evidence for a surprising degree of retention of acquired performance, with absolutely no forgetting evident over retention intervals up to three years in some cases. On the basis of these studies and in agreement with the theoretical position put forth by Kolers and Roediger in their influential article on "procedures of mind," we propose that to understand long-term skill retention, it is useful to assume that memory representations cannot be divorced from the procedures which were used to acquire them, and that the durability of memory depends critically on the extent to which these learning procedures are reinstated at test. According to this argument, tasks like target detection, mental multiplication, and data entry, all of which require the direct storage and retrieval of specific procedures, should be acquired and maintained with much greater facility than tasks which involve procedural memory more indirectly, such as the standard list learning experiments, even those involving short-term recall. The primary reason for making this assumption is that in the traditional verbal learning experiments, the memory coding procedures used by subjects to store the list are not easily retrieved or reinstated at the time of test, unless the subjects employ specific mnemonic procedures, such as the method of loci, the keyword method, or the chunking method used by Ericsson and Chase's expert S.F. In contrast, the procedures used by subjects in our three durable tasks during acquisition are easily and naturally reinstated during the retention test because the subjects are performing the same operations in both cases.

This organizing framework also derives support from the fact that it is consistent with the theory of transfer appropriate processing proposed by Bransford, Franks, Morris, and Stein and the encoding specificity principle espoused by Tulving and Thomson, both of which postulate that memory performance will be best when the test requirements at the time of retention match the encoding requirements employed during learning. Although our framework is

compatible with these earlier approaches, our work differs in four primary respects: First, we examine retention over much longer time intervals (up to three years) than those usually studied by these investigators (typically up to one week maximum). Second, we study tasks that emphasize the acquisition of skill, not merely episodic or declarative aspects. Third, we can more clearly identify and control procedural components of these tasks. Finally, we are able to use the principle of procedural reinstatement to account for permastore, a concept not addressed by other researchers because serious amounts of forgetting were always evident in their studies.

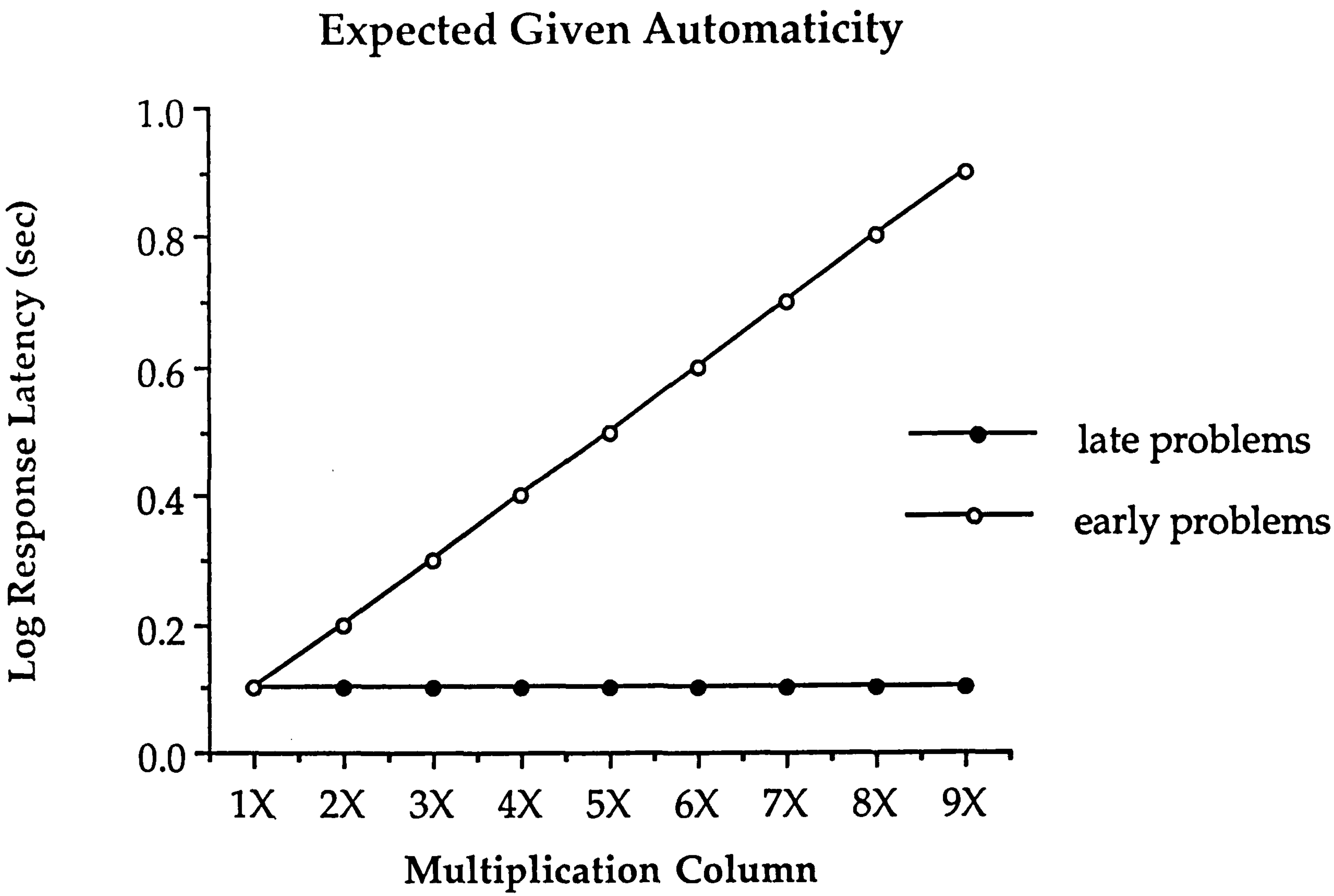
In this presentation, I will first review two of our studies which provide support for this theoretical framework. One of these studies revealed a surprising degree of long-term skill retention, whereas the other yielded considerable forgetting over even relatively short retention intervals. Both of these investigations involve the domain of arithmetic calculation. Next I will outline three studies that we have recently initiated in order to provide more direct support for our theoretical framework. The methods used in these new studies are derived from those used in the earlier ones that I will review. The first of the new investigations also involves the domain of arithmetic calculation, the second introduces the new domain of Morse Code reception, and the third studies the domain of vocabulary learning, which is a domain we have also studied extensively under the structural approach, as Anders will review later.

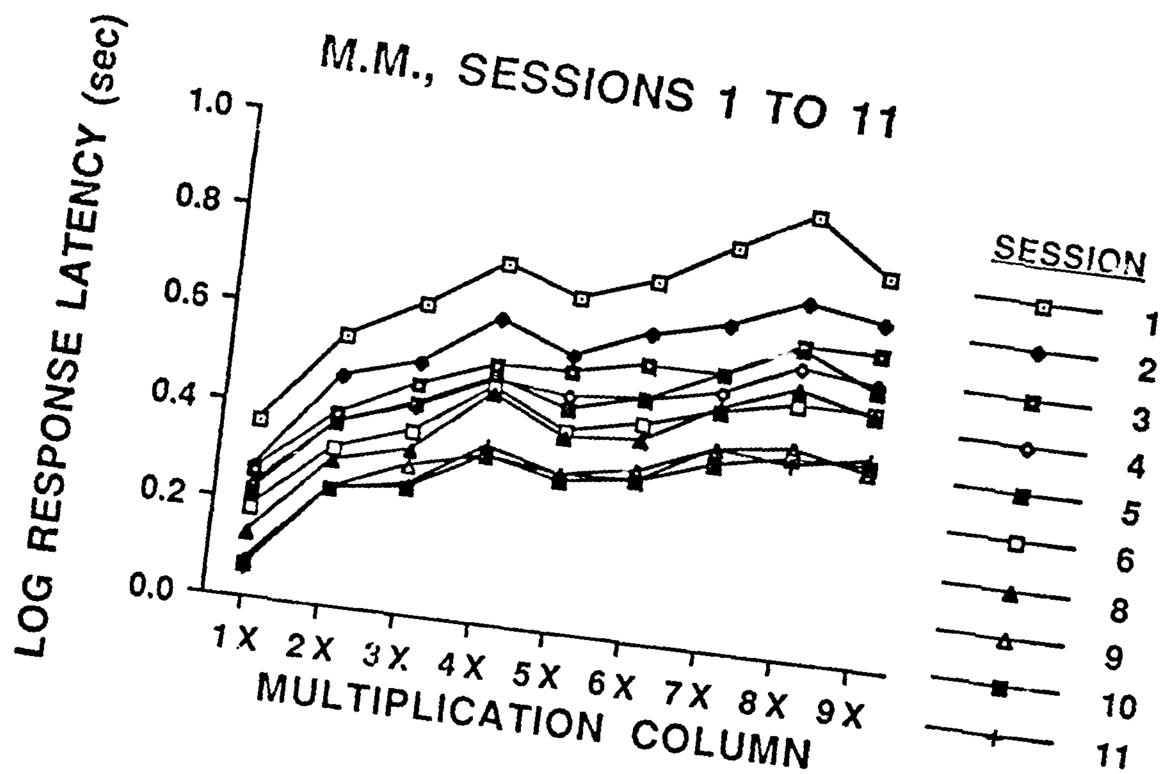
In work conducted in collaboration with David Fendrich, we have found remarkable retention in a paradigm involving training in mental multiplication. In this task as well as others we have studied, we attempt to determine whether retention is related to automaticity. Subjects are shown single-digit multiplication problems, like 3 X 5, and they respond with the answers, either by typing them into the computer or saying them aloud into a microphone. This is a natural task which subjects learned initially outside of the laboratory.

But with additional training subjects show considerable improvement, at least in terms of speed of responding (accuracy is in some cases on the ceiling). I will begin by presenting the data from two subjects given extensive training and tested after substantial retention intervals. Both of these subjects were given 11 training sessions with the typing response and a final training session with the oral response. The subjects were then retested at retention intervals up to 14 months, and each retention test involved the oral response. On each training and testing session the subjects were shown all 81 problems with single digit operands. Individuals typically respond more slowly as the size of the operands increases. For example, responses are typically slower to 8 X 9 than to 2 X 3. Hence, we use as our index of automaticity the function relating the speed of responding to the size of the operands, the problem size effect. As shown in the idealized functions on the first slide (see SLIDE 1), if subjects become automatic, the multiplication column function should flatten, so that there is little or no effect of problem size on response latency.

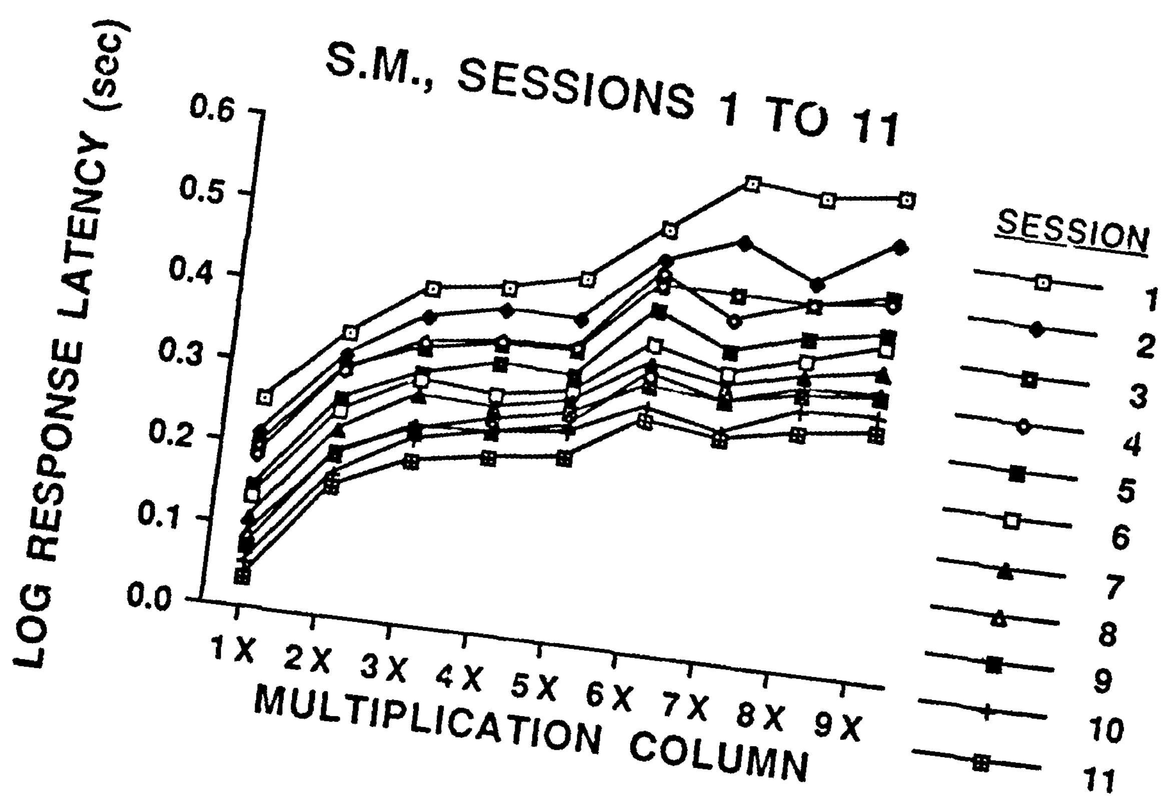
The next slide (see SLIDE 2) shows the acquisition functions for the two subjects. Response latencies are shown as a function of training session number and the multiplication column. It is clear from this figure that both subjects showed large problem size effects, large decreases in latency as sessions increased, but essentially no change in the effect of multiplication column with practice. Hence, the subjects improved at this skill, but they did not become more automatic by our criterion. The next slide (see SLIDE 3) shows the retention data. Specifically, the slide includes functions for the last day of training and each of the retention tests. Note that these subjects show essentially no forgetting, despite the fact that they were not automatic by our criterion.

In two follow-up experiments completed this year our goal was to gain a better understanding of what subjects learn when they are given training in the mental multiplication task. In the first of these experiments we tried to

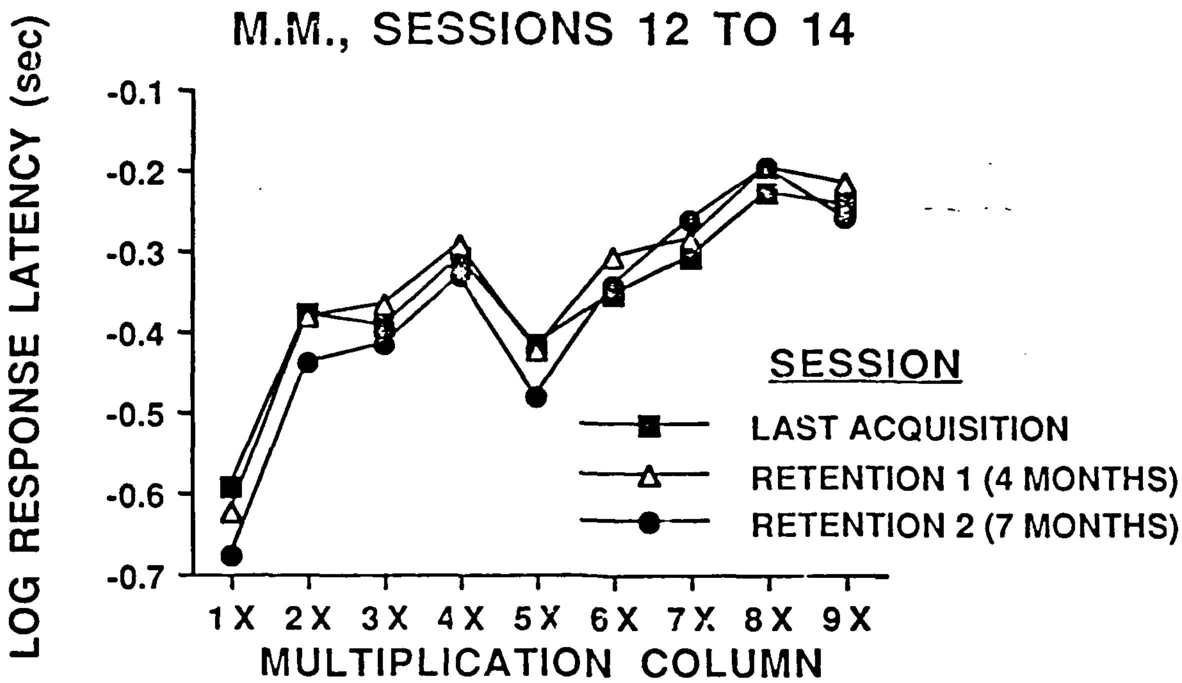

Expected Given Automaticity
Log Response Latency (sec)
1.0
0.8
0.6
0.4
0.2
0.0
1X
2X
3X
4X
5X
6X
7X
8X
9X
Multiplication Column
late problems
early problems

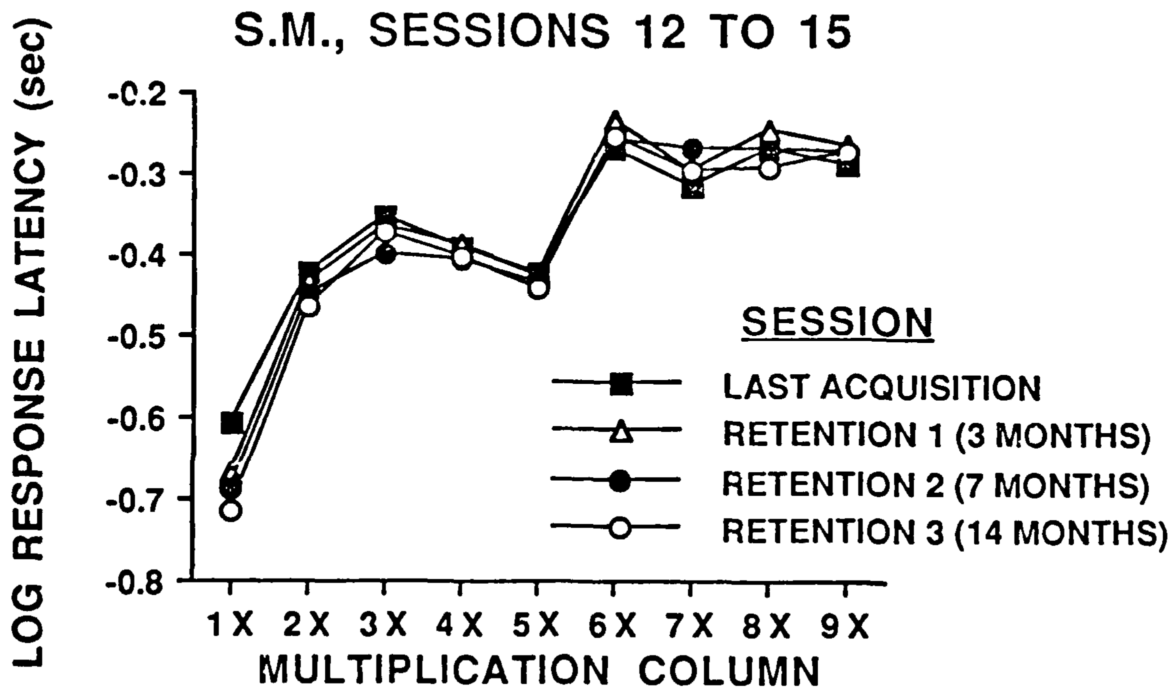


M.M., SESSIONS 1 TO 11
LOG RESPONSE LATENCY (sec)
1.0
0.8
0.6
0.4
0.2
0.0
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
1
2
3
4
5
6
8
9
10
11



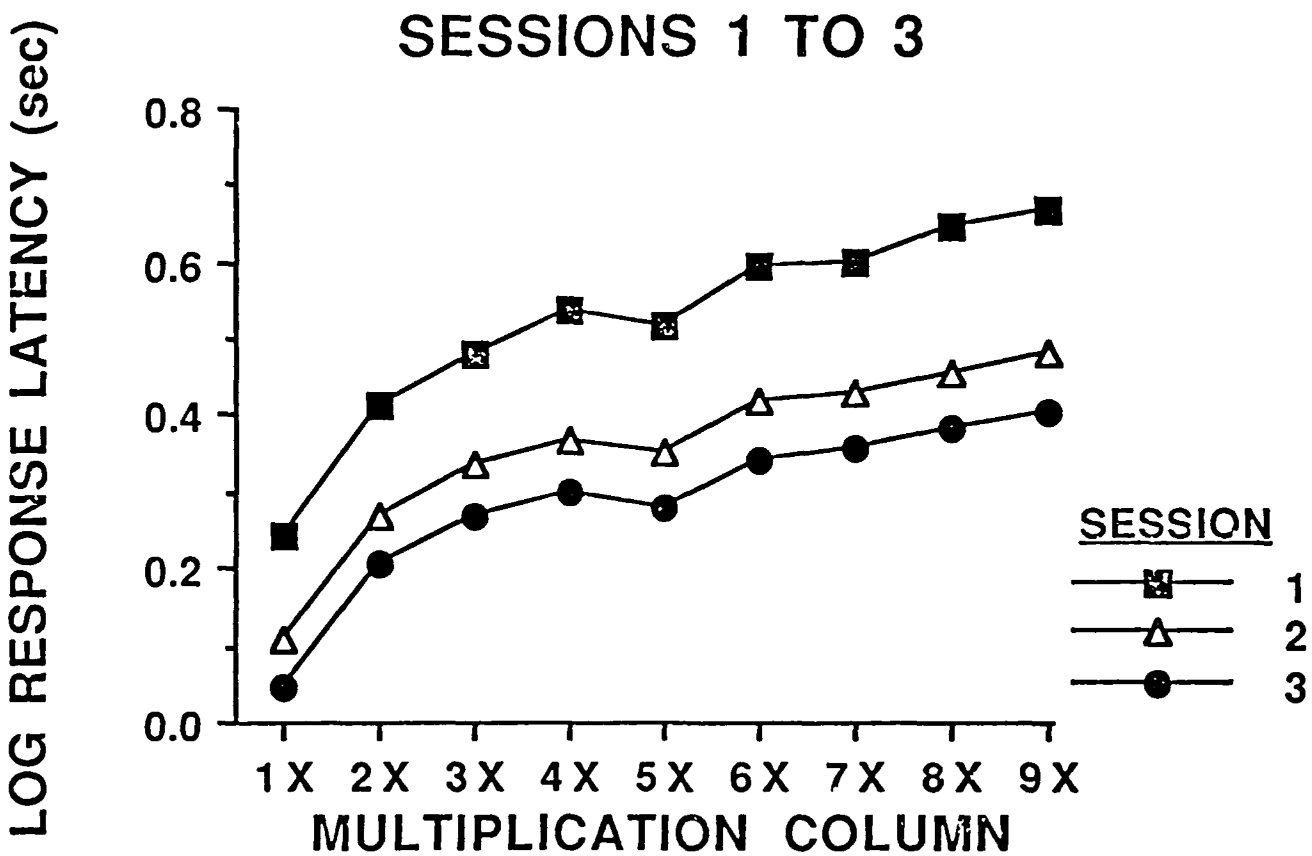
S.M., SESSIONS 1 TO 11
LOG RESPONSE LATENCY (sec)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
1
2
3
4
5
6
7
8
9
10
11

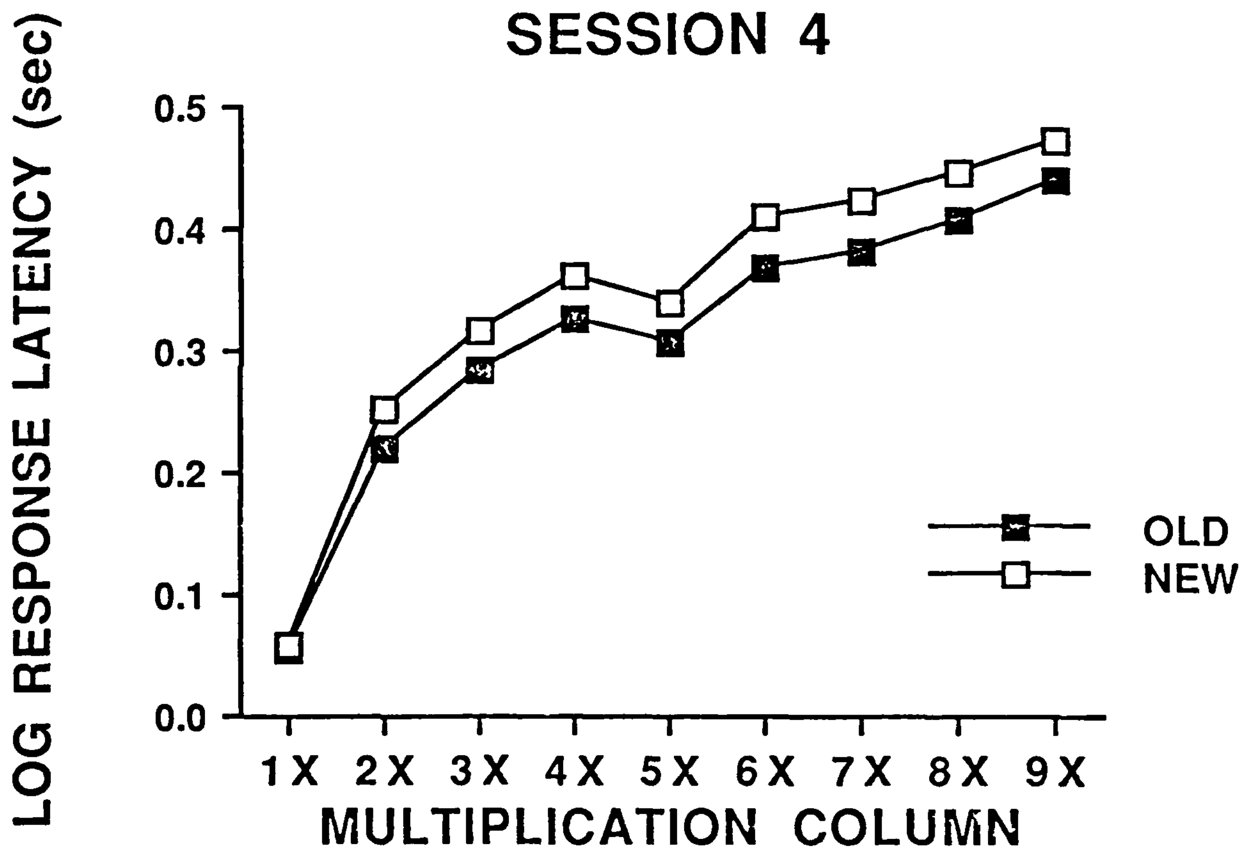

M.M., SESSIONS 12 TO 14
LOG RESPONSE LATENCY (sec)
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
LAST ACQUISITION
RETENTION 1 (4 MONTHS)
RETENTION 2 (7 MONTHS)


S.M., SESSIONS 12 TO 15
LOG RESPONSE LATENCY (sec)
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
LAST ACQUISITION
RETENTION 1 (3 MONTHS)
RETENTION 2 (7 MONTHS)
RETENTION 3 (14 MONTHS)

determine how specific was the information learned by the subjects. In particular, we wondered whether subjects simply strengthened the correct answers and the associations between each answer and the two operands that produce it or whether instead the multiplication operations themselves were strengthened. In order to explore this issue, subjects in this experiment were given training on only half of the multiplication problems. Specifically, during training subjects were shown multiplication problems with single-digit operands. Square problems, such as 2 X 2, were excluded. The remaining 72 problems were divided into pairs, with the two problems in a pair differing only in operand order (for example, 6 X 5 and 5 X 6). Two subsets of problems were constructed with one problem from each pair in each subset. In each of three acquisition sessions subjects were shown problems from one of the two subsets depending on their counterbalancing group. A retention test occurred one month later during which all subjects were shown the complete set of 72 problems. During all four sessions subjects responded by typing their answers using the numeric keypad on the terminal. The next slide (see SLIDE 4) summarizes the acquisition latencies as a function of multiplication column and session. This slide reveals the typical problem-size effect found in earlier studies including our first experiment with subjects given extensive training. Also, as in that previous experiment, response latencies declined as training progressed.

The next slide (see SLIDE 5) summarizes the retention latencies as a function of multiplication column and whether the problem was old (shown during the training phase) or new (not shown earlier). The slide reveals a consistent advantage for the old relative to the new problems across problem size with the single exception of problems in the 1X multiplication column. Presumably the lack of an old/new difference for the 1X problems is due to the fact that subjects do not truly compute the answer to these problem but rather use a simple rule -- namely, if one of the operands is one, the answer is the other operand. Because the new problems differed from the old ones only in operand

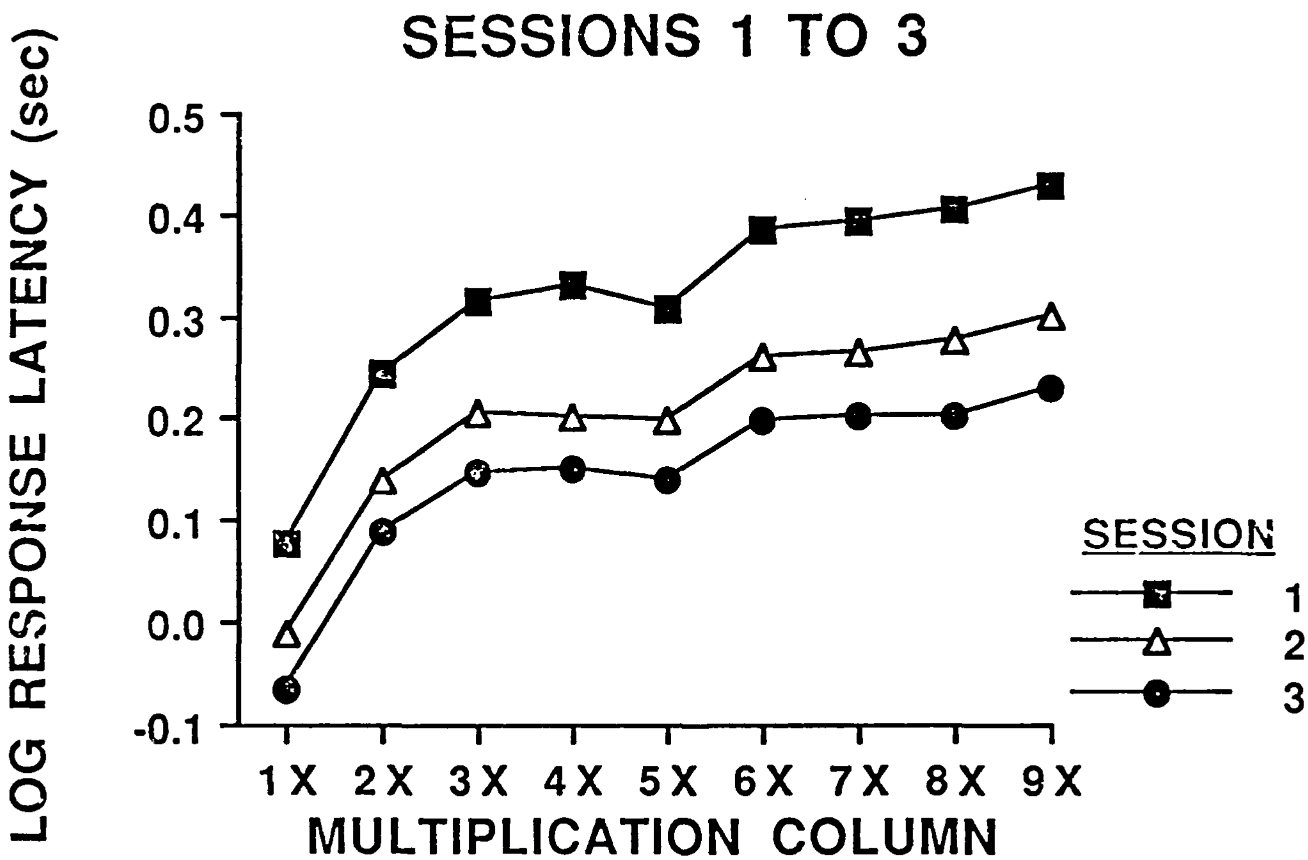

SESSIONS 1 TO 3
LOG RESPONSE LATENCY (sec)
0.8
0.6
0.4
0.2
0.0
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
1
2
3

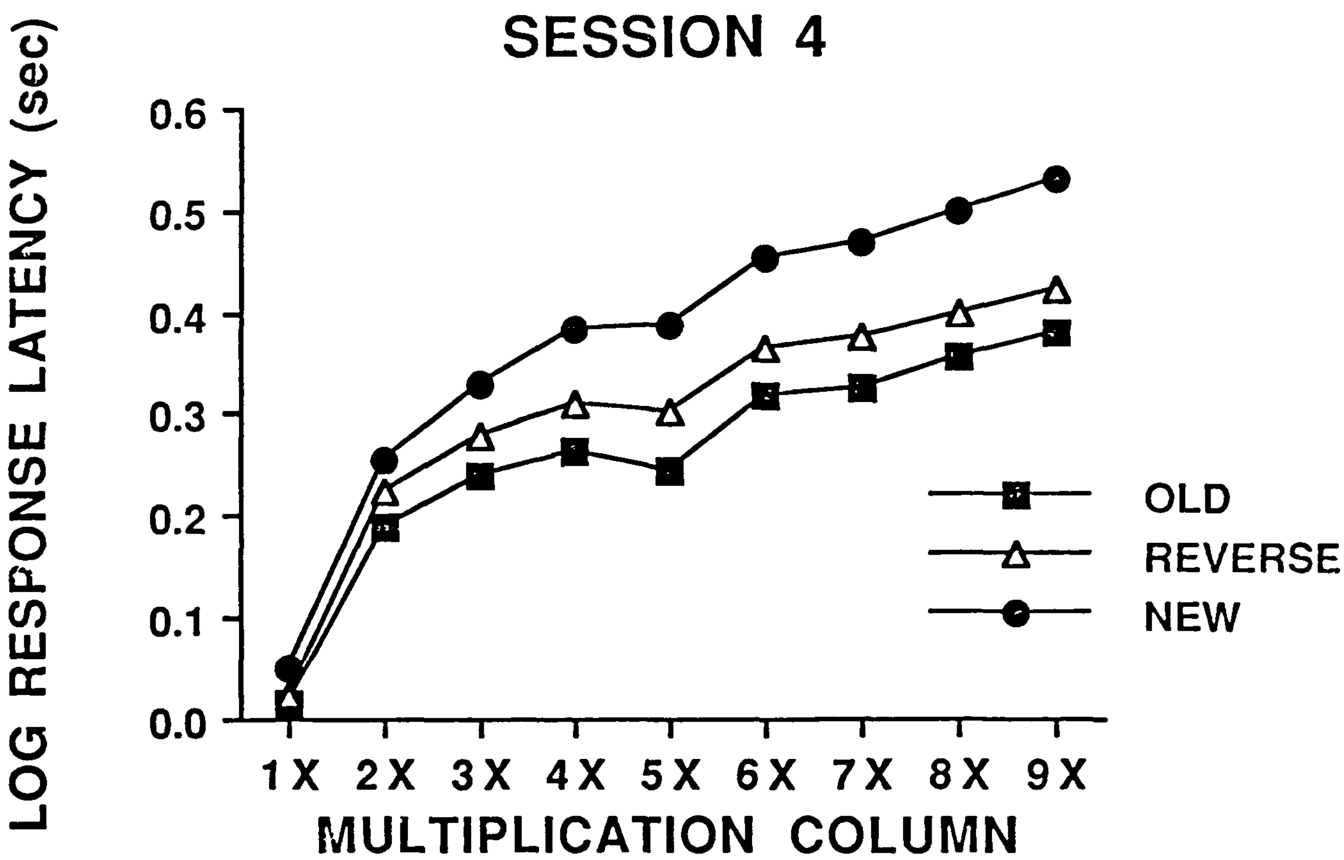

SESSION 4
LOG RESPONSE LATENCY (sec)
0.5
0.4
0.3
0.2
0.1
0.0
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
OLD
NEW

order, the old/new difference found for all but the 1X problems suggests that the information learned by the subjects during training was very specific and concerned the multiplication operations themselves, not just the correct answers or the associations between each answer and the two operands that produce it.

Although this experiment revealed that the new problems with operands in the reverse order were responded to more slowly than the old problems, it is not clear from this study whether there was any facilitation for these new problems due to the practice with the matching problems which had similar multiplication operations. In our second follow-up study we addressed that question. The design was similar to that used in the last experiment except that during acquisition subjects were shown a smaller subset of problems. Instead of pairs, the problems were divided into quadruples, with the four problems in each quadruple including two pairs with problems differing only in operand order. For some of the quadruples, the two pairs had the same answer (for example, 2 X 6, 6 X 2, 3 X 4, 4 X 3). For the remaining quadruples, the two pairs were matched for difficulty as closely as possible. Four subsets of problems were constructed with one problem from each quadruple in each subset. During each of three acquisition sessions subjects were shown problems from one of the four subsets. During the retention session one month later all subjects were shown the complete set of problems.

The next slide (see SLIDE 6) summarizes the acquisition latencies. As in the last experiment, the typical problem-size effect is maintained, although latencies decline, across the three sessions. The following slide (see SLIDE 7) summarizes the retention latencies as a function of the multiplication column and problem type. There are three types of problems in this experiment: old, reverse, and new. The reverse problems were identical to the old ones except that the order of the operands was reversed; these problems had been classified as "new" in the previous experiment. New problems in the present experiment were ones that contained a new combination of operands. All three types of


SESSIONS 1 TO 3
LOG RESPONSE LATENCY (sec)
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
1
2
3


SESSION 4
LOG RESPONSE LATENCY (sec)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
OLD
REVERSE
NEW

problems showed the expected effect of problem size. There was also a consistent advantage for the old relative to the other two types of problems and for the reverse relative to the new problems for all problem sizes except those in the 1X multiplication column, as in the previous experiment. The difference between the reverse and new problems in the present experiment was significant even when considering only those quadruples in which the new problems had the same answers as the old and reverse problems. This finding indicates that practice on problems transfers to those with similar multiplication operations, thereby lending further support to the hypothesis that the information learned by the subjects during practice concerned the multiplication operations themselves.

In our work on memory for numerical calculations, which was conducted in collaboration with Robert Crutcher, we provided a more general test of the importance to memory for procedures or mental operations. This work followed from the phenomenon known as the generation effect. A growing number of experiments have demonstrated a distinct retention advantage for material that is generated by an individual rather than simply read. For example, in these experiments the stimuli are often pairs of words presented under two conditions: <u>read</u> and <u>generate</u>. Sample problems are shown on the next slide (see SLIDE 8). In the <u>read</u> condition, a pair of words is shown, and subjects read the pair aloud (for example, <u>hot</u>, <u>cold</u>). In the <u>generate</u> condition, a word pair is shown with the first word intact and the second missing one or more letters; subjects must then generate the second word of the pair using the first word as a context (for example, <u>white</u>, <u>bla</u>--). If it is assumed that the generation effect is due to the internal activation of auxiliary cognitive operations, then a task leading the subjects to perform such cognitive operations but not necessarily overt generation of an item may show equivalent retention to a generate task. Likewise, a task involving overt generation by the subjects but no auxiliary cognitive operations may not result in any better retention than a read task.

Sample Trials

Read condition

subject sees	subject says
hot, cold	hot, cold

Generate condition

subject sees	subject says
white, bla_ _	white, black

In other words, according to this formulation it is not essential that the subjects generate or produce the stimulus, but rather it is essential that the subjects engage in the auxiliary cognitive operations, or mental procedures, linking the stimulus to other information stored in memory.

In order to test this cognitive operations hypothesis, we devised an experimental paradigm which allows for the orthogonal variation of stimulus presence (absent or present) and auxiliary cognitive operations (self or other). Four tasks are included, which we call the "read," "generate," "verify," and "calculate" tasks. Subjects in all four tasks are given single-digit multiplication problems. As shown on the next slide (see SLIDE 9), which presents a sample problem for each task, in the read task the answers are present in the problems and the multiplication operations are performed by another agent (the experimenter), whereas in the generate task the answers are absent and the multiplication operations are performed by the subjects themselves. The verify and calculate tasks are the ones crucial for testing the cognitive operations hypothesis. In the verify task the subjects are given a problem with its answer but are required to verify that the answer is correct. Contrastingly, in the calculate task, the subjects must provide the answers to the problems but they are told to use a calculator rather than perform the arithmetic themselves. After completing all the problems, subjects in this study were asked to recall the answers to all the problems they had been shown. The cognitive operations hypothesis yields the prediction that retention on the verify and generate tasks would be superior to that on the read and calculate tasks, because in the former two tasks the multiplication operations are performed by the subjects themselves whereas in the latter two tasks the multiplication operations are performed by another agent (either the experimenter or a calculator). In contrast, no difference is expected between the generate and verify tasks or between the calculate and read tasks, because the difference between whether the answers are absent or present in the problems

Task	Subject Sees	Calculator Available	Subject Responds
Read	6 x 8 = 48	No	"6 x 8 = 48"
Generate	6 x 8 = ?	No	"6 x 8 = 48"
Calculate	6 x 8 = ?	Yes	"6 x 8 = 48"
Verify	6 x 8 = 48	No	"6 x 8 = 48, correct"

is not thought to be of much consequence. The results of primary interest are summarized on the next slide (see SLIDE 10) in terms of proportions of correct responses on the free recall test. In accordance with our hypothesis, recall was greatly affected by whether or not the subjects performed the mental operations but not by whether they were shown the answers with the problems.

In a follow-up experiment we aimed to assess the generalizability of these results. More specifically, our goal was to replicate and extend the findings from our first experiment along two dimensions. First, we sought to determine whether the same pattern of results would be obtained for retention over considerably longer delays than were involved in the immediate testing situation of the first experiment. Second, we aimed to assess whether a recognition test procedure would lead to the same findings as the recall procedure used in the first experiment.

The method was similar to that in the first experiment except that subjects were tested either immediately, after a two-day delay, or after a seven-day delay. Right after the recall task subjects were given the recognition test. Subjects were shown pairs of multiplication products and for each pair they were to circle the one number in the pair that was an answer to one of the multiplication problems they were given during the study phase.

The results of the recall task are summarized on the next slide (see SLIDE 11) in terms of proportions of correct recall responses for the four tasks in each of the three retention interval conditions. As in Experiment 1, recall levels for the generate and verify conditions were higher than those for the read and calculate conditions, and this same pattern of results was found for each of the three retention interval conditions although increased delay between study and test did depress performance levels considerably.

The results of the forced-choice recognition task are summarized on the next slide (see SLIDE 12) in terms of proportions of correct recognition responses. Although performance levels for the recognition task were higher

	Cognitive Operations	
Stimulus Presence	Self	Other
	Verify	Read
Present	.68	.38
	Generate	Calculate
Absent	.68	.42

Stimulus Presence and Retention Interval	Cognitive Operations: Self	Cognitive Operations: Other
Present	Verify	Read
Immediate	.59	.42
Two-day	.40	.24
Seven-day	.24	.10
Mean	.41	.25
Absent	Generate	Calculate
Immediate	.55	.34
Two-day	.49	.16
Seven-day	.40	.14
Mean	.48	.21

Stimulus Presence and Retention Interval	Cognitive Operations Self	Other
Present	Verify	Read
Immediate	.82	.81
Two-day	.76	.61
Seven-day	.72	.52
Mean	.77	.65
Absent	Generate	Calculate
Immediate	.81	.65
Two-day	.75	.66
Seven-day	.69	.64
Mean	.75	.65

than for the recall task, the same pattern of results was found for recognition as for recall. Specifically, recognition levels were higher for shorter delays between study and test and, most crucially, were higher for the generate and verify conditions than for the read and calculate conditions, with essentially no differences between the generate and verify or between the read and calculate conditions.

It is important to note that although performance was influenced by the use of cognitive operations or mental procedures in this task, in no case was performance at the ceiling, and we did find substantial decreases in performance over retention intervals up to a week, so that permastore contributes little to these results.

I will now review three of our new experiments which we have initiated to provide more direct support for our theoretical framework. Recall that in the work on memory for numerical calculations which I just reviewed, we studied the generation effect as it applied to subjects' memory for simple multiplication problems and answers shown to them during an experimental session. In brief, we found support for procedures theory by showing that memory was improved if subjects performed the multiplication procedures themselves instead of simply reading the problem with the answer provided by the experimenter or using a calculator to derive the answer. In other work on mental multiplication, we studied the effects of training on the speed and accuracy with which subjects performed mental multiplication. We found, for example, that subjects showed significant improvement as training progressed and that this improvement was retained over a 14-month retention interval, despite the fact that performance had not yet reached the level of automaticity. In the first of our new studies, which is being conducted in collaboration with Danielle McNamara, we have combined these two areas to see whether the generation effect and the support for the procedural account will apply to learning to solve multiplication problems (that is, semantic memory, or memory for facts) as opposed to memory

for the specific problems encountered during the experimental session (that is, episodic memory, or memory for autobiographical events).

Undergraduate students are receiving three sessions of training on simple multiplication problems like that given in our earlier work on mental multiplication. Three different types of training are being compared. Sample problems for each condition are shown on the next slide (see SLIDE 13). The read condition involves subjects reading and copying the problems and answers which are shown on the computer terminal. The generate condition involves subjects reading and copying the problems (but not the answers) and then computing and typing the answers to the problems. And the calculate condition involves subjects reading and copying the problems and then waiting for the computer to calculate and display the answer, which the subject then types. More specifically, for example, subjects in the read condition are shown the equation "2 x 3 = 6" and they type "2 x 3 = 6", whereas subjects in the generate condition are shown "2 x 3 = " and they type "2 x 3 = 6." Finally, subjects in the calculate condition are shown "2 x 3 = ", and they type "2 x 3 = ". Next the computer displays the answer "6", which the subject then types. We expect that all three groups of subjects will show improvements in typing speed and accuracy as training progresses. Before training begins, at the end of training, and approximately one month after training, subjects are given a multiplication test in which they see problems like those studied earlier and must provide the answers by typing them into the computer. The testing schedule for the experiment is outlined on the next slide (see SLIDE 14).

We predict that subjects in the generate condition will be faster and more accurate at the end of training and on the retention test than subjects in the read and calculate conditions, in accordance with our earlier work. Such a finding would be of interest for at least four reasons. First, we would have direct support for our procedural account of long-term retention. Second, the generation effect would be extended to semantic memory. Third, there would be

Sample Trials

Read training

computer	subject
2 * 3 = 6	2 * 3 = 6

Generate training

computer	subject
2 * 3 =	2 * 3 = 6

Calculate training

computer	subject	computer	subject
2 * 3 =	2 * 3 =	6	6

Testing

(Pretest, Post-test, Retention test)

computer	subject
2 * 3 =	6

answer: __

Testing Schedule

Week 1

Monday....................................	Pretest Training-10 blocks
Wednesday................................	Training-10 blocks
Friday......................................	Training-10 blocks Posttest

Week 5

Friday......................................	Retention test

practical implications for the conduct of refresher training of skills learned earlier. Specifically, such results would imply that trainees should be required to perform the relevant mental operations themselves rather than simply be shown the results of the operations performed by another agent. Fourth, there would be educational implications concerning the use of calculators by those individuals who are trying to improve their mathematical skills. Specifically, our findings would imply that it would be best to discourage individuals from relying on a calculator or other computation aid when learning and performing computations.

Although subjects in the generate condition are expected to do better than those in the read and calculate conditions after training, subjects in the two latter conditions may show some improvement relative to the initial test taken before training begins. This result would imply that reading and typing or using a calculator without engaging in the mental multiplication operations are effective, although not optimal, methods of refresher training. More generally, our findings as a whole might suggest that the best way to retain factual information would be to practice retrieving it but that simply reading or copying the information may provide some minimal benefits.

Our second experiment moves us from the domain of mental arithmetic to the domain of Morse Code reception. At the ARI contractors' meeting last year in Ft. Gordon, Georgia, we met Dr. Robert Wisher, an ARI researcher who introduced us to an intriguing problem of individual differences involving the training of receivers of Morse Code. Although solving this problem does not fit in with our immediate plans, we would like to address the related issue of the long-term retention of the skill learned by the Morse Code receivers. Our initial experiment on this topic, which is being conducted in collaboration with Deborah Clawson, would enable us to extend the generality of our theoretical framework to a skill which is in current use by the Army trainees and, unlike mental multiplication, does not involve arithmetic calculations. We have begun

training undergraduate students on Morse Code reception. In this first study we are not using training procedures identical to those employed at the Morse Code Intercept Training program at Ft. Devens; rather, we are employing procedures as closely analogous as possible to those used in our related experiments involving arithmetic calculations. In future research, we expect to make our training closer to that actually used by the Army. More specifically, as shown on the next slide (see SLIDE 15), in our initial study subjects receiving training hear sequences of long and short tones, each sequence corresponding to a different alphanumeric signal in Morse Code (e.g., dah, dah, dah). Subjects hit the appropriate keys to echo this sequence of tones and then either generate or read (calculation is not a reasonable option in this case) the corresponding alphanumeric character (e.g., O), which they type on the keyboard of a computer terminal. During an initial pretest, a posttest after training, and a retention test one month later, subjects hear a Morse code signal and respond simply by typing the corresponding alphanumeric character. The testing schdule for this study is like that for the previous experiment on mental multiplication training. The schedule is shown on the following slide (see SLIDE 16). As for the previous experiment, we expect that at the end of training and after a one-month delay, responses will be faster and more accurate for subjects trained in the generate condition than for those trained in the read condition. Also, we predict that for the read group as well as for the generate group, performance at the end of training and on the retention test will be better than on the initial test taken before training begins. Not only will this study help confirm our findings in a new domain of more direct relevance to the military, it will also give us experience with and insight into this skill. Further, this initial investigation may throw some light on the interesting problem of individual differences in the acquisition of Morse code reception. For example, an examination of individual differences among subjects in the read condition might reveal that trainees differ in the extent to which they naturally use the

Sample Trials

Generate training

computer	subject
dah dah dah	dah dah dah, O

Read training

computer	subject
dah dah dah, O	dah dah dah, O

Testing

(Pretest, Posttest, Retention test)

computer	subject
dah dah dah	O

Testing Schedule

Week 1

Monday.................................... Pretest
Training-4 blocks

Wednesday................................. Training-20 blocks

Friday..................................... Training-20 blocks
Posttest

Week 5

Friday..................................... Retention test

generate strategy, and this difference may be correlated with their training success.

We have recently completed the pilot work for our third new experiment. This experiment involves the domain of vocabulary learning. In our pilot study, which was conducted by Danielle McNamara, eight subjects were given three days of training in which they learned foreign language equivalents of English nouns. A retention test followed one week after the last day of training. The testing schedule for this experiment is shown on the next slide (see SLIDE 17). All subjects were presented with the same 30 word pairs, which were composed of 30 English words, all of which were one-syllable common concrete nouns, paired with 30 nonwords, all of which were single nonsense syllables that had been elicited by a separate group of subjects. We used nonsense syllables rather than real words in a foreign vocabulary so that we could be sure that our subjects had no previous experience with the foreign language and we could control the relation between the English and foreign words. Specifically, the pairing of English words and nonsense syllables was constrained so that the two words in a pair did not rhyme and there were no obvious mnemonic relationships between them.

Subjects were initially provided with the list of the 30 word pairs and allowed ten minutes to familiarize themselves with the list. They were then given a cued-recall pretest in which they were provided with the English words of each pair and asked to recall as many of the corresponding foreign words as possible. At this point the subjects were divided into two training groups, read and generate. Both groups completed a total of 15 blocks of training over the three acquisition sessions, with each block containing all of the 30 word pairs. As shown on the next slide (see SLIDE 18), subjects in the read group were first shown the English word in a pair which they copied on a sheet of paper. Next the foreign word equivalent was displayed, which the subjects copied beside the English word. Subjects in the generate group also copied the English word in a pair but then generated the corresponding foreign word. These

Testing Schedule

Week 1

Wednesday	Pretest Training
Thursday	Training
Friday	Training Posttest

Week 2

Friday	Retention test

Sample Trials

Read training

Experimenter	subject	Experimenter	subject
light	light	yord	yord

Generate training

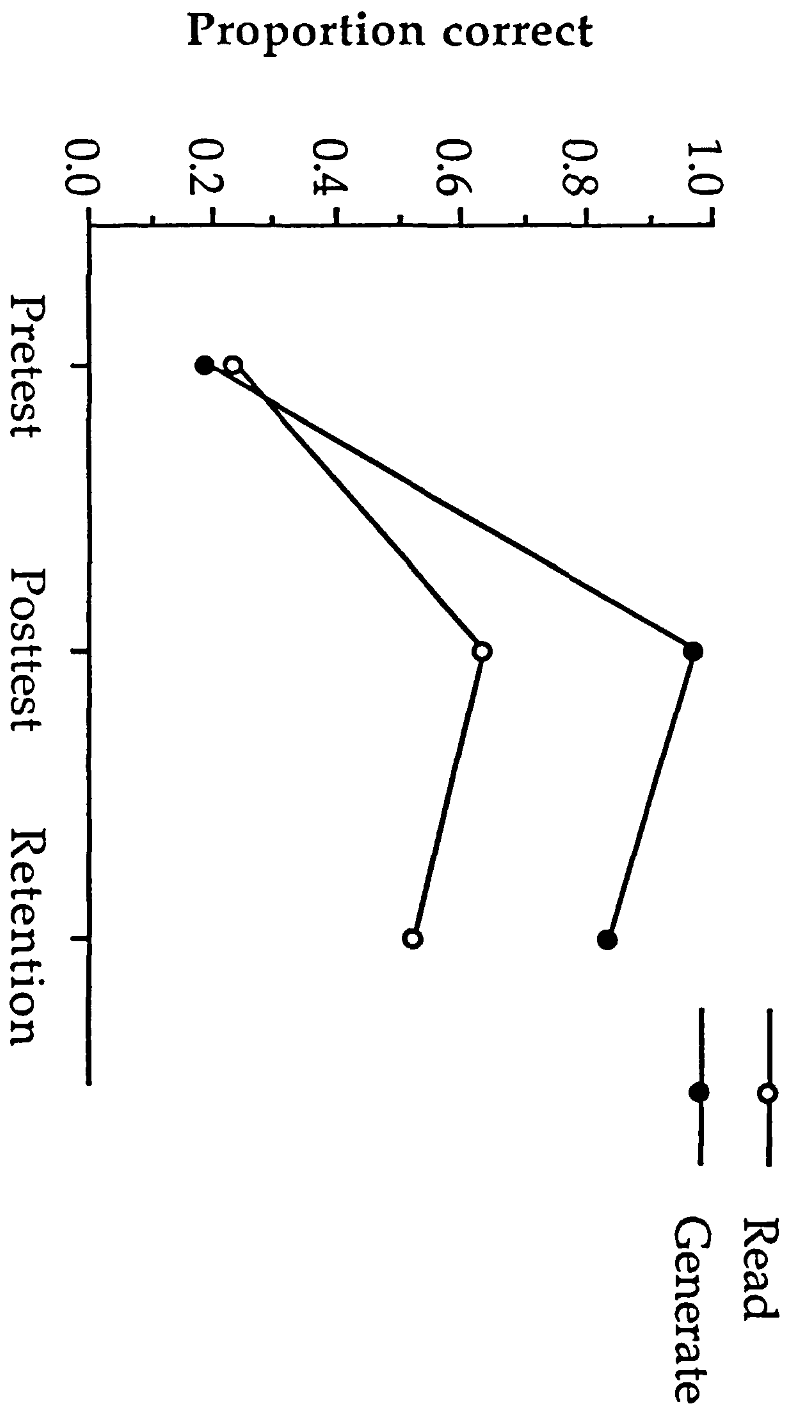
Experimenter	subject	Experimenter
light	light yord	yord

subjects were shown the correct foreign word only after they had written down the English word and their response. If their response was incorrect, the subjects copied the correct foreign word equivalent of the English word. In a posttest administered at the end of the last acquisition session and on a retention test given one week later, subjects from both the read and generate groups were administered a cued recall test identical to that given in the pretest. At the conclusion of the retention test subjects were asked to provide retrospective verbal reports about the mnemonic strategies, if any, they had used to memorize the word pairs.

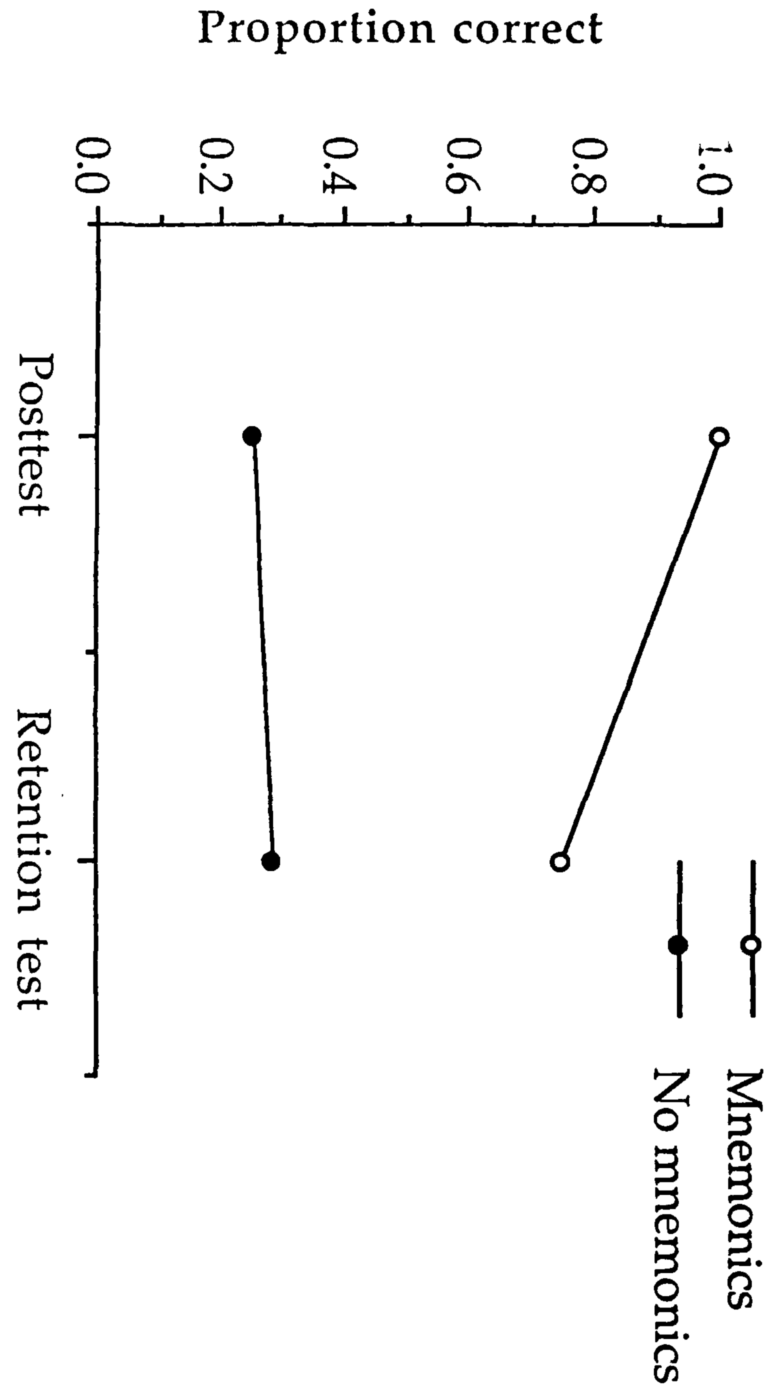
The results of the pretest, posttest, and retention tests for the read and generate groups are summarized on the next slide (see SLIDE 19) in terms of the proportions of correct recall responses. These results indicate clearly that the process of generating the nonwords during the training sessions had a significant positive effect on the proportion of words correctly recalled at both the posttest and the retention test one week later.

It is further interesting to note that whether or not the subjects used mnemonic strategies to memorize the words seemed to have an important relation with acquisition and retention performance. Two subjects in the read group reported using few or no mnemonic strategies to memorize the words. As shown on the next slide (see SLIDE 20), these subjects had low scores on the posttest and retention test. In contrast, the remaining two subjects in the read group (like all four subjects in the generate group) reported using numerous mnemonic devices, and these subjects recalled significantly more words on both the posttest and the retention test.

Unlike previous studies reported in the literature on the generation effect, we found a large difference between the read and generate conditions in this pilot experiment even though the recalled items were nonwords and different groups of subjects were exposed to the read and generate conditions. The generation effect found here along with the difference between the subjects in



Proportion correct
0.0
0.2
0.4
0.6
0.8
1.0
Pretest
Posttest
Retention
Read
Generate


Proportion correct
0.0
0.2
0.4
0.6
0.8
1.0
Posttest
Retention test
Mnemonics
No mnemonics

the *read* group who used mnemonic strategies and those who did not point to the importance of employing mnemonic procedures for enhancing retention. The subjects who exhibited higher recall of the vocabulary items clearly developed procedures, that is mnemonic strategies, which were used to retain the word pairs. The results of the analysis of the *read* group suggest that the adoption of such useful mnemonic procedures does not depend on the experimental condition, although the *generate* condition seems to promote such procedures. We argue that these procedural processes constitute the critical factor determining long-term retention, and that the generation effect is simply a reflection of the subjects' tendency given particular experimental conditions to create procedural links between cues and learned responses.

In conclusion, although the theoretical framework we have proposed, which is centered around the notion of procedural reinstatement, is able to throw considerable light on our understanding of the long-term maintenance of knowledge and skills, it can by no means account for all important retention phenomena. Other theoretical constructs are needed to provide a complete account of long-term retention, and the constructs we have outlined need to be fleshed out in greater detail and require more rigorous experimental tests that also determine their generalizability to other tasks and situations. Nevertheless, we have been impressed with the wide variety of memory studies that already fit into this framework and the remarkably large range of forgetting rates found in these studies. If procedural reinstatement can indeed explain these large differences in forgetting rates, as we have argued, the practical significance of this finding is substantial. The implication is clear that if we wish to retain knowledge or skills over a long delay interval, it is crucial that we make sure that the procedures we use when learning the information are reinstated at the time we need to recall the information.

The goal of the structural approach is to assess the detailed structure of acquired knowledge and skill. Based on this detailed analysis of the cognitive processes, Rob Crutcher and I hope to uncover distinctions between rapidly decaying and more stable knowledge and skill components. Our primary focus has been on the retention of foreign vocabulary items in Spanish, following up the work by Bahrick which has shown permanent retention of a proportion of the knowledge originally acquired during instruction in Spanish. We will examine the accuracy of producing correct translations after different delay intervals and also, in line with the earlier reported studies, we will examine the retention of speed of retrieval for correct responses at these delays.

Based on our review of the literature on the acquisition of foreign language vocabulary as well as our own pilot studies, three rather distinct types of foreign vocabulary items have been identified. They are shown in the following slide.

Insert Slide 1 about here

For the first type of vocabulary items, the English word is perceptually similar to the foreign word, which makes memorization easy. For the second type, there exists a perceptually similar word in English, which is semantically related to the target word. The third type is characterized by an absence of any relation between the two words.

For the third type, which presents students with the greatest difficulty, investigators have advocated a number of different mnemonic techniques to learn

Three Types of Forgeign Vocabulary Items

Perceptually related

limen	lemon
palaco	palace
valle	valley

Semantically related

periodico	journal	(periodical)
angulo	corner	(angle)
serpiente	snake	(serpent)

No relation

doronico	leopard
sonrisa	smile
cabra	goat

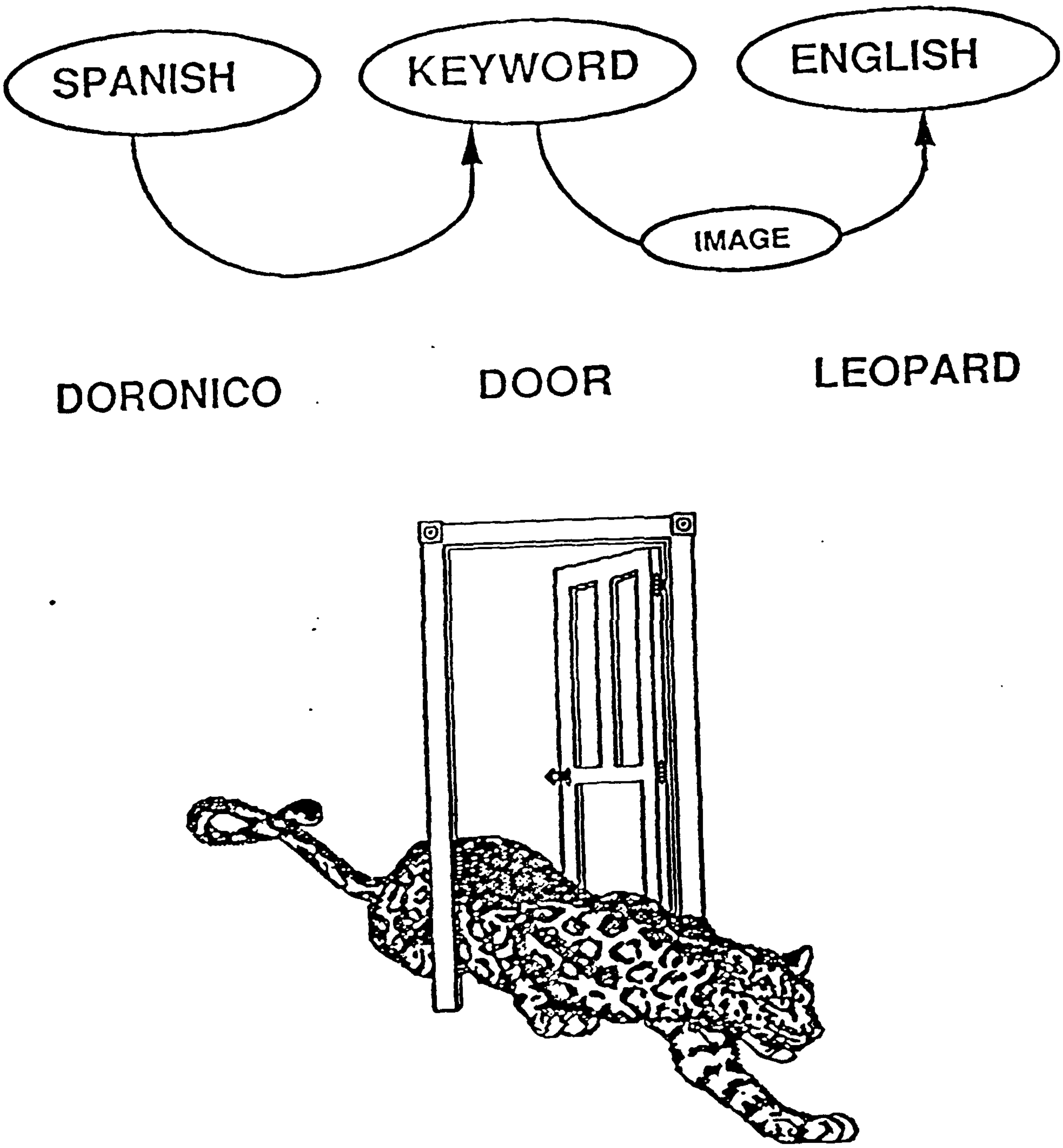
the items. The most popular and successful of these techniques is the keyword mnemonic. In the keyword mnemonic, subjects are given an English word which is perceptually similar to the Spanish word and then asked to form an interactive image linking the two English words, as illustrated in the next slide.

Insert Slide 2 about here

The Spanish word doronico accesses the English word door (called the keyword), which in turn is linked by an interactive image between door and the English translation -- leopard. Let us now go back to the three types of vocabulary items and propose a schematic representation for how they might be encoded in memory, which is illustrated in the next slide.

Insert Slide 3 about here

The memory representation of all three types of items contains a perceptual relation between the Spanish word and an English word. The last two types contain an additional link necessary to access the English target word. In our current work we are comparing the retention of all three types of items, but in our initial work we focused on learning of the third type of vocabulary items by using the earlier described keyword method.



Example vocabulary item for keyword method.

palaco $\longrightarrow$ palace

angulo $\longrightarrow$ angle $\longrightarrow$ corner

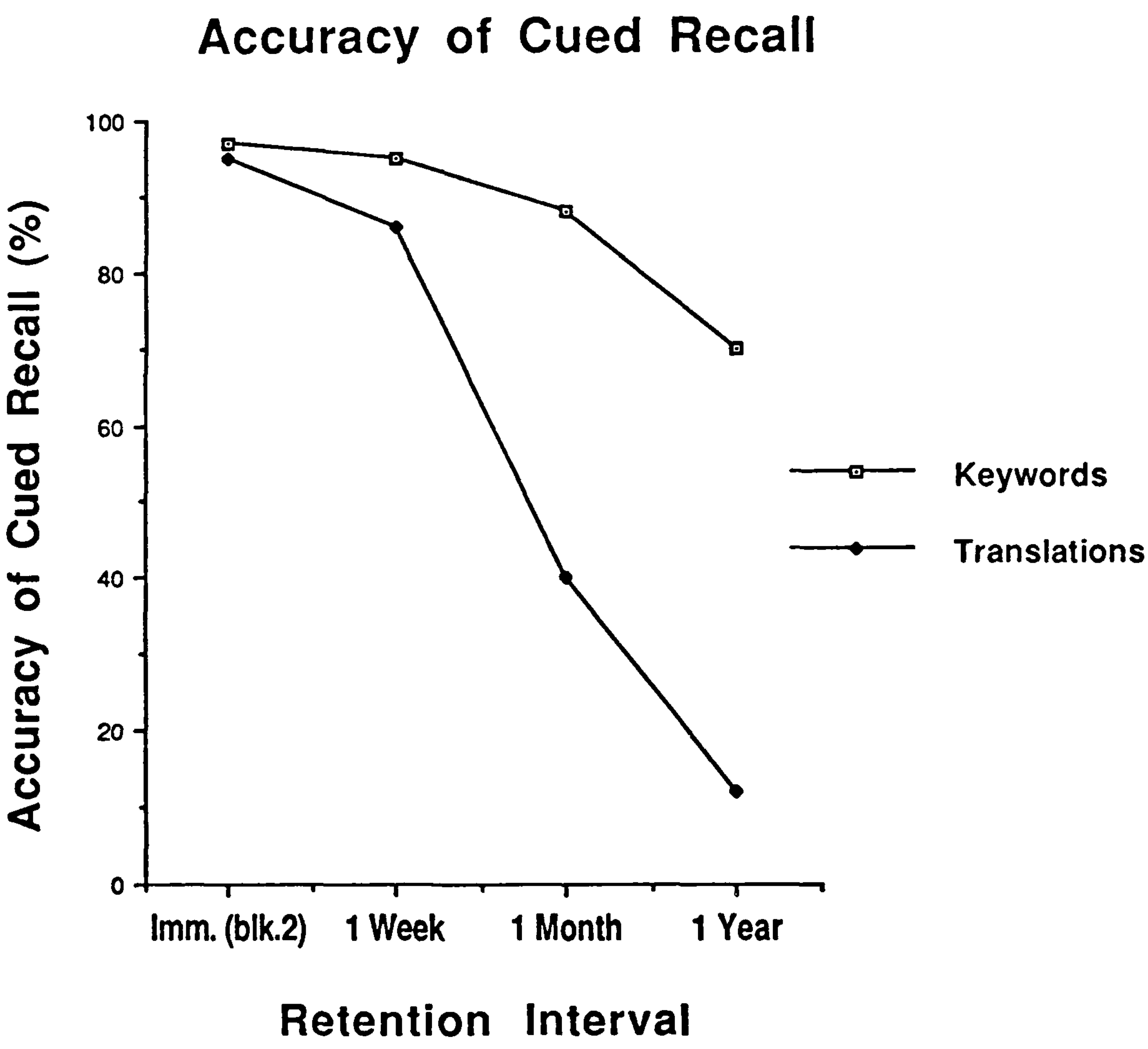
doronico $\longrightarrow$ door $\longrightarrow$ leopard

In addition to examining the retention of the English translation of each of the Spanish words, we have also examined retention of the keywords by instructing subjects to generate the keyword corresponding to each Spanish word at immediate and delayed tests. We will argue that the retention of keywords follows a retention function similar to that for the vocabulary items of the first type. In addition, we examined subject's recall of the English translation, when cued with the corresponding keyword. In our tests we collect data on how well the acquired relations are retained by evaluating the accuracy of cued recall, which is comparable to Bahrick's analysis of retained knowledge. For the correctly recalled items we also measure and analyze the speed of the retrieval process. The analysis of retrieval speed as a function of delay is closely related to the previous reported analysis of maintained retrieval speed in the mental multiplication studies reported earlier by Alice. We also collect retrospective reports on the retrieval process for half of the items.

In the next slide the percent correct recall is graphed as a function of test delay for cued recall of translations and cued recall of keywords. The results for a one-year delay are preliminary and based on a small number of subjects.

Insert Slide 4 about here

The forgetting on the two tasks follows distinctly different patterns, with relatively rapid forgetting of English translations and slow forgetting of the keywords.


Accuracy of Cued Recall
Accuracy of Cued Recall (%)
100
80
60
40
20
0
Imm. (blk.2)
1 Week
1 Month
1 Year
Retention Interval
Keywords
Translations

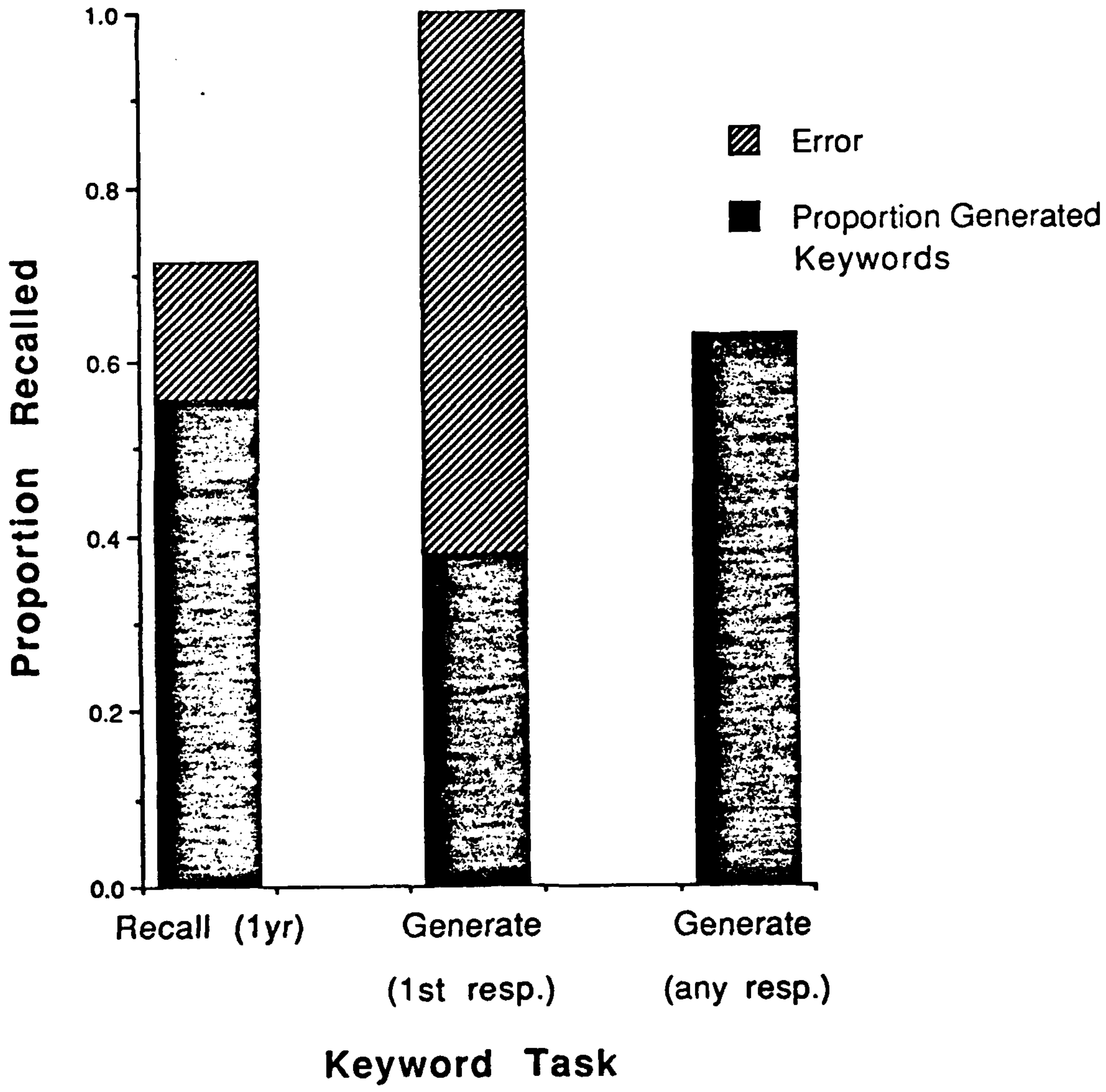
Given the similarity between the Spanish words and the keywords, it is possible that subjects simply generated correct guesses of the keywords as opposed to remembering them. To examine this possibility, we had a group of naive college students generate possible English keywords for the Spanish words. The next slide shows the probability of generating the keyword as the first response, as well as the probability of generating the keyword at all during a 40-second interval, compared to the probability of recalled the keyword one year after learning the items. Given that some of the cued-recall items involved the presentation of keywords as cues for the translations, the results for the one-year retention subjects are based only on trials prior to the presentation of these keywords.

Insert Slide 5 about here

The one-year retention subjects are more accurate than the control subjects in generating keywords as their first response. The evidence for retention is even clearer if we adjust for incorrect responses, also shown in the same slide. At the same time, this analysis raises the possibility that subjects in our retention condition may retrieve their responses through a recognition-and-test procedure as opposed to a direct recall based on the Spanish word as a cue (check verbal reports for suporting evidence).

It is also possible to examine retention by studying the speed of correct retrievals. The next slide shows the average speed of correct retrievals for the first test as a function of delay for both the English translations and the English keywords.

Proportion of Generated Keywords as a Function of Task



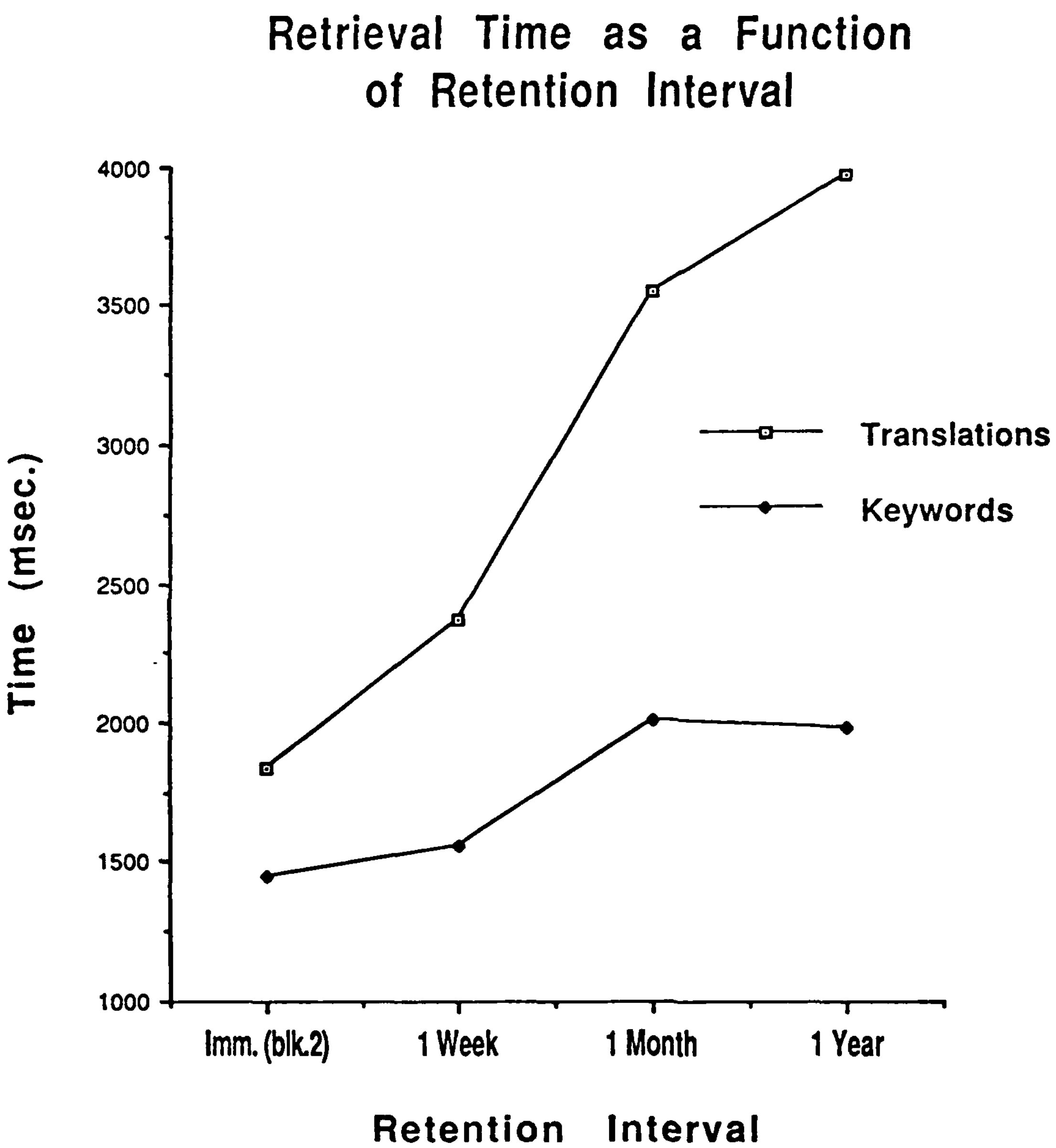
Insert Slide 6 about here

With increased delay we see a longer average reaction time for both types of responses. This pattern of result would at first glance seem completely inconsistent with the results that show virtually perfect retention of acquired speed in multiplication that Alice described earlier. However, when we examine the retrieval latencies for the second time the same items are recalled at delay, we see virtually perfect retention, as shown in the next slide.

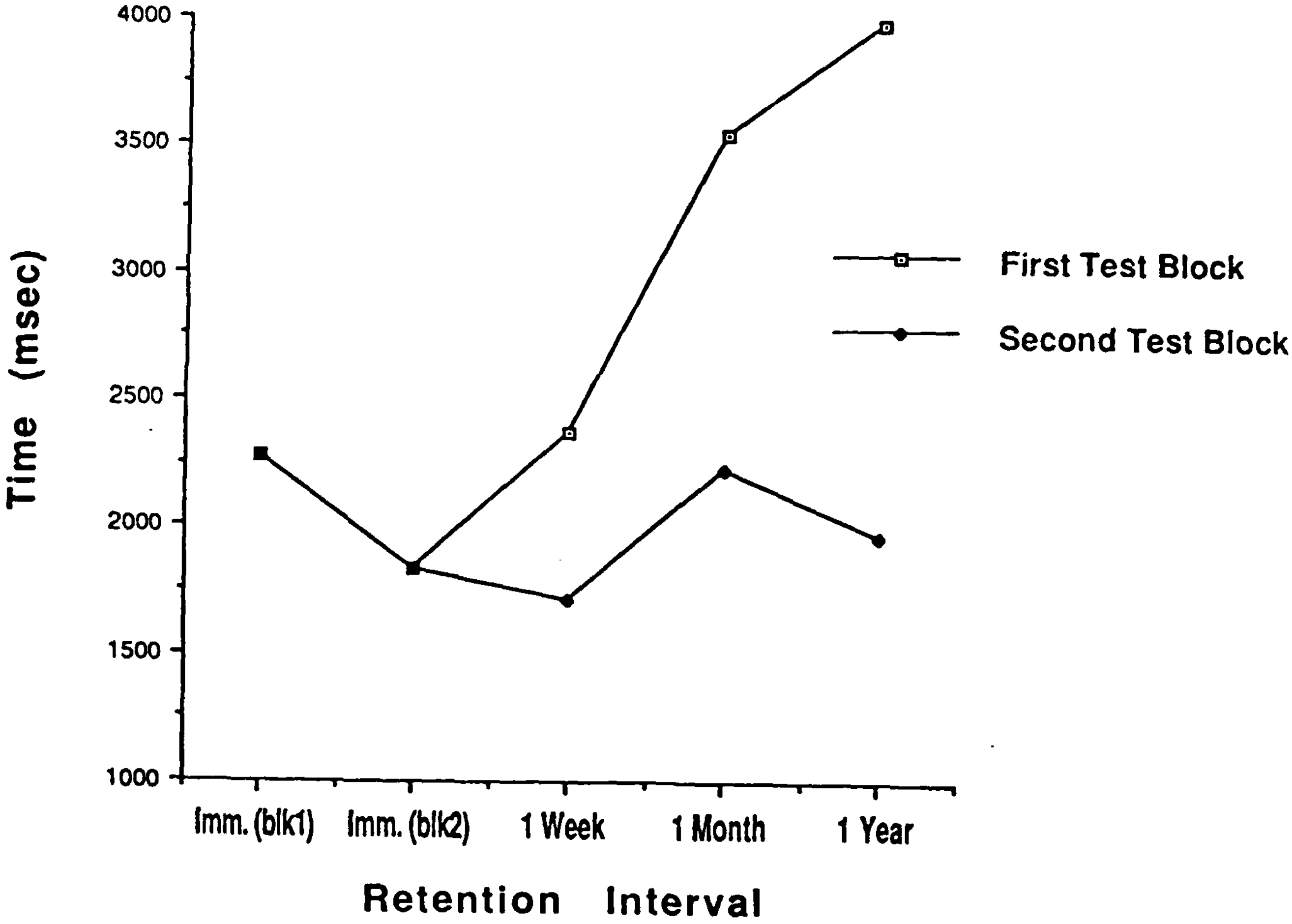
Insert Slide 7 about here

There is a clear improvement in retrieval latency from the first to the second immediate test, and the retrieval latencies for the second retention test appear almost independent of the length of retention interval and are comparable to that of the last immediate test, consistent with the previously reported results.

The observed difference in retrieval speed on the first trial and subsequent trials appear to be quite general and is found in research using the analytic approach and also for well-preserved motor skills, such as typing. This difference raises a number of intriguing theoretical and methodological issues and has important practical implications for retention in tasks where immediate action is required after considerable delay, such as medical emergency procedures like CPR.


Retrieval Time as a Function
of Retention Interval
4000
3500
3000
2500
2000
1500
1000
Time (msec.)
Translations
Keywords
Imm. (blk.2)
1 Week
1 Month
1 Year
Retention Interval

Retrieval Time for Translations as a Function of Retention Interval



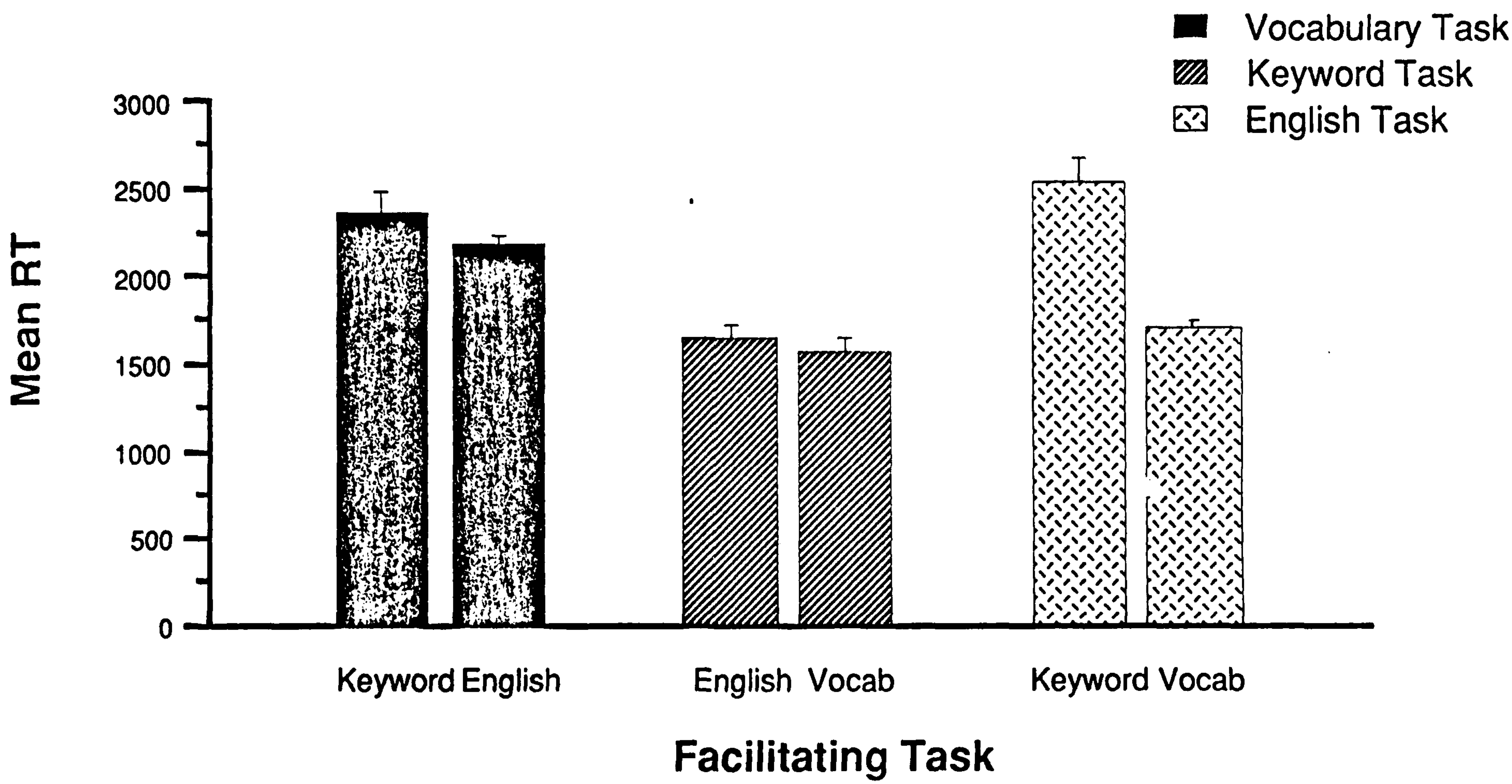
The first retrieval of an acquired piece of knowledge after delay appears to reflect a rather different retrieval process than subsequent retrievals. In support of this claim we found consistent order effects in the first test block at one week's delay. In this test block subjects wre asked to perform three different retrieval tasks for the same items in counterbalanced order.

Insert Slide 8 about here

The next slide shows the retrieval times for the three tasks as a function of which retrieval task was performed on the item on the absolute first test regarding that item. For the task of retrieving the translation from the keyword (image task), subjects were over 500 milliseconds faster when the subjects had already retrieved the translation from the Spanish word (vocabulary task) and thus retrieved the image linking the keyword to the translation compared to cases when the subjects had retrieved the keyword from the Spanish work (keyword task). For the other retrieval tasks no similar order effects were observed.

More generally, we have found strong converging evidence that the retention of vocabulary items is based on retrieval of the keyword as a mediator for retrieval of the English translation from our detailed analysis of RT's, accuracies and verbal reports.

When subjects are able to recall the English translation of a Spanish word at delay they are nearly always able to correctly perform the component retrieval task that is to retrieve the keyword and its association to the English translation.


Vocabulary Task
Keyword Task
English Task
Mean RT
3000
2500
2000
1500
1000
500
0
Keyword English
English Vocab
Keyword Vocab
Facilitating Task

The speed of retrieval on the complete vocabulary task is considerably slower than any of the component retrieval tasks, both at delay and at immediate test.

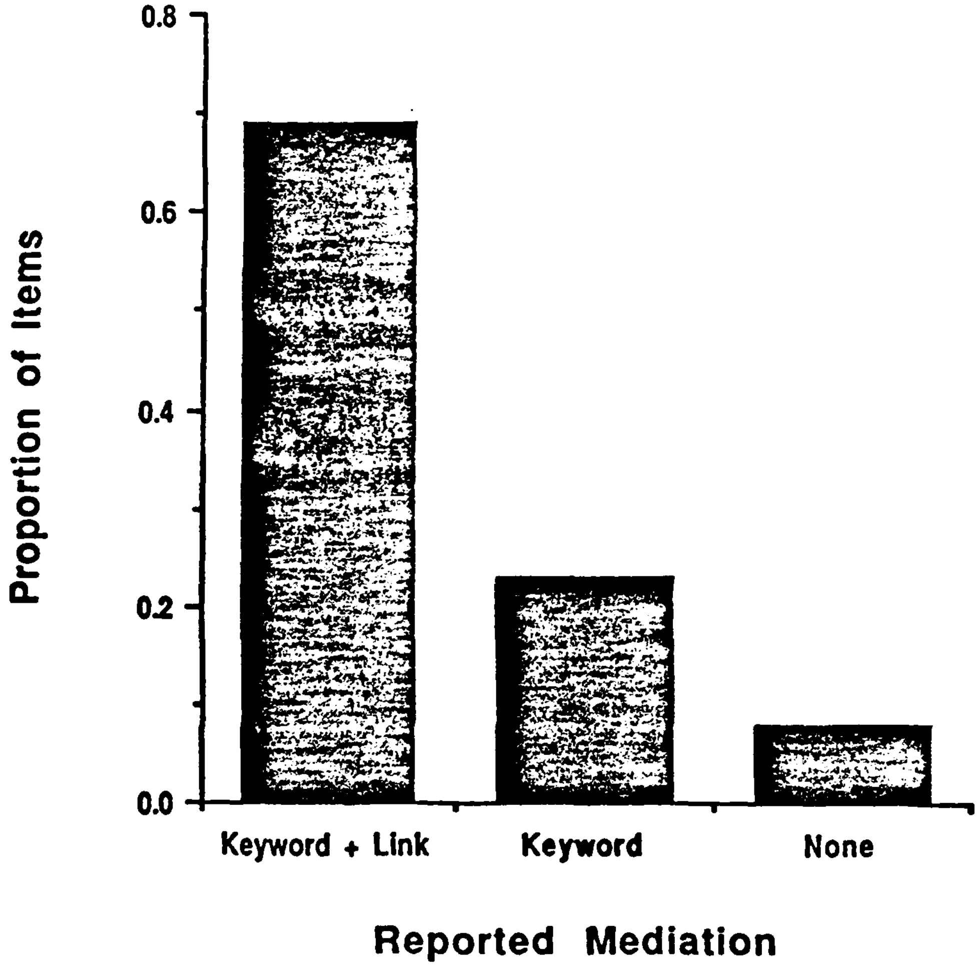
The retrospective verbal reports show that on most of the trials retrieval of the English translations involves access of the mediating keyword and often the image linking the English keyword and the English translation. The frequencies of retrospective reports involving direct/unmediated retrieval, retrieval of only the keyword and retrieval of the keyword and the image are shown for correct retrievals of the English translation in the next slide.

Insert Slide 9 about here

By an analysis comparing the average RTs for trials, where subjects report that retrieval is mediated by access of an image involving the keyword, to RTs for trials without reports of any mediation, we find as would be expected that subjects require a lot more time---on the average 700 milliseconds more when mediation is reported.

In a subsequent experiment we explored the effects of training subjects on only the complete vocabulary task for items versus training only the component retrieval tasks for items. Our basic finding was that training on complete vocabulary tasks improves retrieval speed on the component retrieval tasks and training on the component retrieval tasks improves the retrieval on the complete vocabulary task. which suggests that, as a first approximation, retrieval of vocabulary knowledge is mediated by the English keyword.

Proportion of Vocabulary Items for Different Categories of Mediation (Experiment 1)



Our analysis of retrospective reports on retrieval processes for items recovered after long retention intervals has revealed the issue of how subjects can distinguish correct responses from other accessed incorrect responses. In any situation, where the incorrectly recalled action has serious consequences, it is not sufficient that the subject is more likely than chance to recall the action, unless the subject can distinguish those instances where their recall is accurate. Training leading to the retention of responses with ability to check accurately the correctness of the response might be quite different from traditional training oriented toward increased proficiency and overall retention.

In sum, our studies have shown differential decay for retention of keywords and of the English translations. The decay of retention of English translations is almost completely due to forgetting of the image linking the keyword and English translation. Retrieval speed for retained items decreases with delay for the first access but is almost completely recovered by the second time the same item is retrieved. In our current work we are now examining retention of all the three types of vocabulary items with the same methodology, and we are in particular interested in the retention of vocabulary items of the second type--the semantically mediated items, which may provide us with additional clues to Bahrick's findings of permanently stored knowledge in memory.

The Long-Term Retention of Skills

Alice F. Healy, David W. Fendrich, Robert J. Crutcher, William T. Wittman,

Antoinette T. Gesi, K. Anders Ericsson, and Lyle E. Bourne, Jr.

University of Colorado, Boulder

Acknowledgments. This research was supported in part by United States Air Force Human Resources Laboratory Contract VE5744-022-001 and United States Army Research Institute Contract MDA903-86-K-0155 to the Institute of Cognitive Science at the University of Colorado and by National Institute of Mental Health Training Grant MH14617-08 to the University of Colorado.

The Long-Term Retention of Skills

For the last four years we have been engaged in a study of the long-term retention of skills. Our research project has had as its overarching goal finding ways to optimize long-term retention, particularly the long-term retention of skilled performance. We started with the assumption that some part of acquired knowledge or skill is, for reasons that remain to be identified, permanent. Bahrick (1984), among others, demonstrated that, while a large part of acquired knowledge is lost rapidly, a significant portion can last a lifetime, even if that knowledge is not intentionally rehearsed or accessed in the meantime. In a theoretically noncommittal way, we adopted Bahrick's concept of a permastore as a fundamental fact of memory and we looked for conditions of training or attributes of learned material that lead to permastore. Indeed, in a number of our lines of investigation, we have found evidence for a surprising degree of retention of acquired performance. We will first provide an overview of three of these investigations, and then we will discuss a general theoretical framework which will help us understand this impressive memory performance. In contrast, in other studies we have conducted, we found considerable forgetting over even relatively short retention intervals. We will next review three of these studies, with an attempt to place them in the same general theoretical framework developed to account for the earlier studies. The aim of this review is to derive indications of the specific factors which facilitate retention. Thus, we will try in this chapter to provide an integrated theoretical account of the many different facets of our research program.

We have divided our investigations into two broad categories, those showing some evidence for permanent retention of acquired performance and those showing evidence instead for substantial forgetting; see Table 1. Although Table 1 suggests a sharp boundary separating these two categories, we do not wish to imply that there is a clear separation or that long-term retention is all or none. But this categorization should highlight the factors most crucial in

Table 1

Studies Demonstrating Remarkable Memory (on Left) and Considerable Forgetting (on Right)

High Retention	*Low Retention*
target detection	memory for numerical calculations
mental multiplication	vocabulary learning
data entry	components of memory for course schedules

facilitating memory over a long interval. Our studies demonstrating remarkable memory use target detection, mental multiplication, and data entry tasks. Our studies demonstrating considerable forgetting assess memory for numerical calculations, for vocabulary, and for components of course schedules.

Studies Demonstrating Remarkable Memory

Target Detection

In our studies of target detection (Healy, Fendrich, & Proctor, 1990), subjects sit at a computer terminal where they see displays containing 16 random characters grouped to resemble three words. Half of the displays contain the letter *H*, which is the target, and half do not. The subjects' task is to press a response button every time they see a target, and both response accuracy and latency are used as dependent variables. A major independent variable is frame size. There are either 2, 4, or 16 letters in the display with the other characters replaced by number signs. A sample display of each frame size is shown in Figure 1. Subjects generally respond less accurately and more slowly as frame size increases (see, e.g., Schneider & Shiffrin, 1977), and the loss of this frame size effect with practice is our index of automaticity. We wondered whether retention of target detection skill would be related to the degree of skill automaticity achieved during training.

In our first experiment examining the retention of the detection skill we employed three groups of subjects. Subjects in the control group received no detection training, those in the limited training group performed 10 blocks of training over two sessions, and those in the extensive training group received 24 blocks over four sessions. A retention interval of three to five weeks then elapsed before subjects returned for the final (retention) session. At that time subjects in all three groups performed five blocks of detection training to evaluate retention of the letter detection skill.

The results of detection training for the limited and extensive training groups are shown as a function of day of training and frame size in Figure 2 for

proportion of hits. For both groups hit rates increased during acquisition. A substantial frame size effect was evident even at the end of training, but the effect was significantly reduced with practice, even in the limited training group. In contrast, although response latencies, shown in Figure 3, decreased during acquisition, they showed no reduction in the frame size effect. Comparisons of the last day of training and the retention test showed no decrease in accuracy or speed for either group, suggesting essentially perfect retention of the detection skill over the delay interval.

The retention data are replotted in Figure 4 where they are compared to the analogous data from the control group, in terms of both proportion of hits and response latencies. Hit rate was significantly better for the extensive training group than for either of the other groups. Also, for response latencies as well as hit rate, the frame size effect was significantly smaller for the extensive training group than for the other two, suggesting that indeed responding did become more automatic after extensive training.

One purpose of our next experiment was to determine whether more intensive practice would lead to more dramatic changes in the frame size effects and, therefore, a greater degree of automaticity. A second purpose was to assess retention of the detection skill over delays considerably longer than the one-month interval used in our first study.

This study employed only two subjects, both of whom were given 12 days of practice on the detection task. Both subjects were recalled 6 months later for a retention test, and one of the subjects (A.G.) was also recalled 9 months after that (i.e., 15 months after the initial training) for a subsequent retention test. The results in terms of accuracy are shown in Figure 5. Large improvements in performance are evident during training, so that before the end of training both subjects have attained essentially perfect accuracy, even with the largest frame size. Most interesting is the observation that this maximal level of performance is maintained across the long retention intervals.

Although no forgetting was evident, some forgetting could have been masked by a ceiling effect because accuracy was so high. This problem does not exist, however, when we examine response latencies, which are shown in Figure 6. Again there are large improvements in performance with practice, but the functions for the three frame sizes do not completely converge in this case, especially for D.S. Thus, the subjects became more automatic with practice but did not achieve full automaticity. Nonetheless, performance on the retention tests showed a remarkable degree of memory. Specifically, after 6 months D.S. showed absolutely no forgetting, and although A.G. did show some loss at that point, her response latencies after the 15-month retention interval were no different than at the end of training.

Mental Multiplication

We also find remarkable retention when we move to a paradigm involving mental multiplication (Fendrich, Healy, & Bourne, 1989). In this task as well, we are examining whether retention is related to automaticity. Subjects are shown single-digit multiplication problems, like 3 X 5, and they respond with the answers, either by typing them into the computer or saying them aloud into a microphone. This is a natural task which subjects learned initially outside of the laboratory, but with training subjects show considerable improvement, at least in terms of speed of responding (accuracy is in some cases on the ceiling). We will first report the data from two subjects given extensive training and tested after substantial retention intervals. Both of these subjects were given eleven training sessions with the typing response and a final training session with the oral response. The subjects were then retested at retention intervals up to 14 months, and each retention test involved the oral response. On each training and testing session the subjects were shown all 81 problems with single-digit operands. Individuals typically respond more slowly as the size of the operands increases. For example, responses are typically slower to 8 X 9 than to 2 X 3. Hence, we use as our index of

automaticity the function relating the speed of responding to the size of the operands. As subjects become automatic, this function should flatten, so that there is little effect of problem difficulty on response latency.

Figure 7 presents the acquisition functions for the two subjects with the typing response method. Response latencies are shown as a function of multiplication column and session. Both subjects showed large effects of multiplication column, large decreases in latency as sessions increased, but essentially no change in the effect of multiplication column with practice. Hence, the subjects improved at this skill, but they did not become more automatic by the above-stated criterion. Figure 8 shows the retention data. Specifically, Figure 8 includes functions for the last day of training and each of the retention tests with the oral method of responding. Note that as with the target detection task, subjects show essentially no forgetting, despite the fact that they were not automatic by our criterion.

In two follow-up experiments our goal was to gain a better understanding of what subjects learn when they are given training in the mental multiplication task. In the first of these experiments we tried to determine how specific was the information learned by the subjects. In particular, we wondered whether subjects simply strengthened the correct answers and the associations between each answer and the two operands that produce it or whether instead the multiplication operations themselves were strengthened. In order to explore this issue, subjects in this experiment were given training on only half of the multiplication problems. Specifically, during training subjects were shown multiplication problems with single-digit operands. Square problems, such as 2 X 2, were excluded. The remaining 72 problems were divided into pairs, with the two problems in a pair differing only in operand order (for example, 6 X 5 and 5 X 6). Two subsets of problems were constructed with one problem from each pair in each subset. In each of three acquisition sessions subjects were shown problems from one of the two subsets depending on their counterbalancing group.

A retention test occurred one month later during which all subjects were shown the complete set of 72 problems. During all four sessions subjects responded by typing their answers using the numeric keypad on the terminal. Figure 9 summarizes the acquisition latencies as a function of multiplication column and session. This figure reveals the typical problem-size effect found in earlier studies including our first experiment. Also, as in that previous experiment, response latencies declined as training progressed.

Figure 10 summarizes the retention latencies as a function of multiplication column and whether the problem was old (shown during the training phase) or new (not shown earlier). The figure reveals a consistent advantage for the old relative to the new problems across problem size with the single exception of problems in the 1X multiplication column. Presumably the lack of an old/new difference for the 1X problems is due to the fact that subjects do not truly compute the answer to these problem but rather use a simple rule -- namely, if one of the operands is one, the answer is the other operand. Because the new problems differed from the old ones only in operand order, the old/new difference found for all but the 1X problems suggests that the information learned by the subjects during training was very specific and concerned the multiplication operations themselves, not just the correct answers or the associations between each answer and the two operands that produce it.

Although this experiment revealed that the new problems with operands in the reverse order were responded to less quickly than the old problems, it is not clear from this study whether there was any facilitation fcr these new problems due to the practice with the matching problems which had similar multiplication operations. In our second follow-up study we addressed that question. The design was similar to that used in the last experiment except that during acquisition subjects were shown a smaller subset of problems. Instead of pairs, the problems were divided into quadruples, with the four problems in each quadruple including two pairs with problems differing only in

operand order. For some of the quadruples, the two pairs had the same answer (for example, 2 X 6, 6 X 2, 3 X 4, 4 X 3). For the remaining quadruples, the two pairs were matched for difficulty as closely as possible. Four subsets of problems were constructed with one problem from each quadruple in each subset. During each of three acquisition sessions subjects were shown problems from one of the four subsets. During the retention session one month later all subjects were shown the complete set of problems.

Figure 11 summarizes the acquisition latencies. As in the last experiment, the typical problem-size effect is maintained, although latencies decline, across the three sessions. Figure 12 summarizes the retention latencies as a function of the multiplication column and problem type. There are three types of problems in this experiment: old, reverse, and new. The reverse problems were identical to the old ones except that the order of the operands was reversed; these problems had been classified as "new" in the previous experiment. New problems in the present experiment were ones that contained a new combination of operands. All three types of problems showed the expected effect of problem size. There was also a consistent advantage for the old relative to the other two types of problems and for the reverse relative to the new problems for all problem sizes except those in the 1X multiplication column, as in the previous experiment. The difference between the reverse and new problems in the present experiment was significant even when considering only those quadruples in which the new problems had the same answers as the old and reverse problems. This finding indicates that practice on problems transfers to those with similar multiplication operations, thereby lending further support to the hypothesis that the information learned and retained by the subjects concerned the multiplication operations themselves.

Data Entry

The third task we studied in which subjects showed remarkable retention is a motor task involving data entry (Fendrich, Healy, & Bourne, 1988). Subjects

were shown lists of digits and typed them with the keypad of a computer terminal. In our first experiment subjects were given three days of training at entering lists of ten three-digit numbers. A given list was repeated either once or five times during training, with the repetitions either spaced or massed. One month later they were given a retention test in which they entered some of the old lists of numbers along with some new lists. We found in this experiment no reliable effects of either the amount or spacing of repetitions, but we did find a significant advantage for the old lists relative to the new ones at the retention test. Response latencies as a function of day of training and test list type are summarized in Figure 13. Subjects improved at the data entry task, and their performance on the old lists was maintained across the month-long retention interval. Although performance at test was worse on the new lists than on the old lists, even with new lists performance was better than at the start of training.

It was most surprising to us that after a one-month interval there was a difference between the old and new lists of digits, despite the fact that subjects were given no instructions to memorize the lists. However, we noted that our index of memory -- namely, changes in response latency -- was an indirect or "implicit" measure, to use the term first proposed by Graf and Schacter (1985). In a follow-up experiment we addressed the question whether subjects under the same circumstances would also exhibit reliable memory for the lists using a direct or "explicit" measure. If not, we would demonstrate a clear independence or dissociation between the two types of memory measures. Alternatively, perhaps subjects would demonstrate significant memory by the explicit as well as the implicit measure of memory, in which case the question arises whether the memory processes underlying these two types of measures interact in any way.

In order to investigate these issues in our second experiment we modified the procedures of the first study to include a recognition test as an explicit

measure of memory. Specifically, subjects were asked to give a recognition rating on a six-point scale for each digit list shown at the retention test. Subjects were not told about the recognition rating until the beginning of the retention test, so that acquisition still involved incidental learning. For half of the subjects, the rating for each list was given immediately after the list was entered on the keypad, and for the other subjects the rating was made before the list was typed.

As in the first study, we found that subjects' typing latencies significantly decreased as training progressed and changed very little over the one-month delay interval, as shown in Figure 14. Also, we found that the latencies on the retention test were significantly faster for the old lists than for the new lists, as shown in Figure 15. Unlike our first experiment, we found a significant difference between the old and new items only for old items that were repeated five times (old5 lists), not for those that were shown only once (old1 lists).

Most crucial is the signal detection analysis of data from the recognition test. The sensitivity statistic we used was da. As shown in Figure 16, subjects' accuracy was significantly greater than chance for the items shown five times, but not for those shown only once. This result indicates that subjects did have significant memory for the digit lists presented one month earlier by our explicit as well as our implicit measure. Hence, it is interesting to determine whether there is any interaction between the memory processes underlying these two different measures. In fact, there was a significant interaction of repetitions by the order of the recognition and entry tasks. Recognition for the repeated items was better when the recognition test came after, rather than before, the subjects entered the numbers. Hence, typing the digit lists aided the subjects in making their recognition decisions. These results suggest that the memory processes reflected by the explicit and implicit measures are not independent but instead mutually support each other.

To explore further the relation between the two measures of memory, we examined the entry latencies contingent on whether or not subjects correctly classified the digit lists on the recognition test, collapsing the rating scale into a binary "old/new" response. As shown in Figure 17, which presents the mean latencies for the new lists and the old lists repeated five times, the difference in typing latencies between old and new digit lists was significant only when those lists were correctly classified, not when subjects made incorrect recognition judgments. This pattern of results further indicates that memory processes underlying the explicit and implicit measures go hand in hand. Evidence for reliable memory by the implicit measure is only available when there is also evidence of reliable memory by the explicit measure.

In a subsequent experiment we sought to determine the locus of our long-term priming effect. In particular, we wondered whether it was reflecting only a motor component of the data entry task or whether a perceptual component was involved as well. In order to separate the motor and perceptual components, in this experiment we made use of the fact that there are two different conventional orientations of the keypad. As shown in Figure 18, one keypad orientation is used on most computer and calculator keyboards, and the other orientation is used on the standard touch-tone telephone. Subjects were trained on one of these keypads and then switched to the other at the retention test one week later. We included new lists of digits as well as two different types of old lists on the final test. The "old digit" lists included the same sequences of digits as shown during training but required new motor responses. In contrast, the "old motor" lists included new sequences of digits but ones that required the same sequence of motor responses. For example, for the sequence 7539 shown during training, the old digit list would also be 7539, but the old motor list would be 1593, which requires the same sequence of finger movements on the alternate keypad. The latencies to initiate the first digit of each sequence are shown in Figure 19 for the new, old digit, and old motor lists on

the retention test. Subjects showed a significant advantage relative to the new lists for both types of old lists, thereby locating the long-term priming effect at both the motoric and perceptual stages of processing.

Possible bases of permastore. We have now discussed three different tasks, all of which show evidence for strong long-term skill retention implicating Bahrick's (1984) notion of permastore. At this point we can ask what these three tasks have in common, so that we can generate an hypothesis concerning the factors responsible for entry into permastore. Let us start by ruling out several hypotheses which are inconsistent with our findings. First, the results from the target detection and mental multiplication studies are clearly inconsistent with our original idea that entry into permastore is necessarily associated with automaticity. In those two tasks we found essentially perfect retention with little or no indication that subjects had achieved automaticity. Second, we can rule out the hypothesis that only implicit memory measures can reveal evidence for permastore. Although changes in response times were our primary indices of learning and retention in our three tasks and such changes are implicit or indirect measures of memory, our study of data entry clearly demonstrated long-term retention using an explicit memory measure as well. In fact the processes underlying the implicit and explicit memory measures were shown to be interdependent in that task. Third, we can rule out the hypothesis that only motor learning yields superior long-term retention. This hypothesis has some support in the early literature comparing verbal and motor learning (see, e.g., Naylor & Briggs, 1961). Indeed, motor learning was implicated to some degree in the data entry experiments, but perceptual information was also shown to be well retained in those studies. Further, motor learning presumably played only a minor role in the largely perceptual target detection task. Most crucial in this regard are the results of the mental multiplication study. Motor learning was eliminated as a contributing factor in that case, because the subjects given extensive training in the first experiment were tested using an

oral response although they were trained with a typing response. Also, the typing response was used in the retention session as well as the training session for the follow-up multiplication experiments. But in those cases the subjects answered the old problems more rapidly than the corresponding reverse problems even though the answers, and hence the motor responses, were the same. The mental multiplication task is probably best described as a cognitive skill rather than a perceptual or motor skill, so perceptual, motor, and cognitive skills can all gain entrance into permastore.

What do the three tasks we studied have in common? It seems to us that the most important common feature shared by these tasks is a major or overriding procedural, as opposed to declarative, component, to use the distinction made by Anderson (1983), among others. In agreement with the theoretical position put forth by Kolers and Roediger (1984), we propose that memory representations cannot be divorced from the procedures which were used to acquire them, and that the durability of memory depends critically on the extent to which the learning procedures are reinstated at test. Tasks like target detection, mental multiplication, and data entry, all of which require the direct storage, retrieval, and use of specific procedures, should, according to this argument, be acquired and maintained with much greater facility than tasks which involve procedural memory more indirectly, and which place a greater emphasis on events, facts or declarative components, such as the standard list learning experiments. In the traditional studies involving list learning, even those tapping short-term memory (see, e.g., Estes, 1972), the memory coding procedures used by subjects to store the list are not easily retrieved or reinstated at the time of test, unless the subjects employ specific mnemonic procedures, such as the method of loci or the chunking method used by Ericsson and Chase's (1982) expert S.F. In contrast, the procedures used by subjects in our three tasks during acquisition are easily reinstated during the retention test because the subjects are performing the same task (for example, letter detection) in both cases.

This characterization of memory is consistent with Morris, Bransford, and Frank's (1977) theory of transfer appropriate processing and Tulving and Thomson's (1973) encoding specificity principle, both of which postulate that memory performance will be best when the retrieval operations required at the retention test match or overlap the encoding operations employed during learning.

One of our experiments on data entry provides direct support in this domain for the importance of transfer appropriate processing or test appropriateness (Gesi, Fendrich, Healy, & Bourne, 1989). In this study subjects were presented with four-digit sequences on a computer screen. Half the sequences were shown only once, and the other half were shown three separate times in the study session. In one condition the subjects simply read each sequence and pressed the space bar once for each digit in the sequence. In the second condition they entered the sequence using the numeric keypad of the terminal, and in the third condition they entered the sequence using the horizontal number row on the console keyboard. One week after the training session, subjects were given a retention test. This test required them to enter old and new sequences using in some cases the row and in other cases the keypad configuration. After entering each sequence the subjects also made an old/new recognition decision. The theory of transfer appropriate processing would predict that subjects' recognition would be most accurate for the sequences entered in the same way at acquisition and at test. In other words, using the row at study and at test or using the keypad at study and at test should be better than either reading at study or using a different configuration at study and at test.

Our initial analyses examined the typing latencies during the study phase. Most interesting in these analyses is the effect of repetitions on latencies. Subjects were faster at entering digit sequences that were repeated ($\underline{M}$=2.333 sec) than those that were only presented once ($\underline{M}$=2.365 sec), showing that subjects had implicit memory for the digit sequences, even though they were only

four digits long and, hence, considerably less complex than the lists of ten three-digit numbers used in our earlier studies of data entry.

The second set of analyses examined typing latencies during the retention test phase of the experiment. Of primary concern is whether subjects exhibited memory for the digit sequences after the one-week retention interval. Indeed, latencies were significantly shorter for the old sequences shown previously during the study phase ($\underline{M}$=3.002 sec) than for the new sequences ($\underline{M}$=3.029 sec). This advantage for the old items occurred for the sequences entered at study with either the keypad or the row, but not for those only read at study, as shown in Figure 20. This production effect is similar to the generation effect found for episodic memory (see, e.g., Slamecka & Graf, 1978) but points to the crucial role of procedural memory. Although there was a strong production effect on latencies, the effect of test appropriateness was only marginally reliable for this implicit measure of memory. Figure 21 compares latencies for sequences entered the same way at study and test, a different way at study and test, or simply read at study. The latencies for these three conditions were statistically different but the difference is attributable mostly to the disadvantage for the read condition.

The third set of analyses was concerned with the explicit recognition data. The da scores were computed for each subject in each condition. Overall recognition was reliably greater than chance, indicating that subjects did have explicit, as well as implicit, memory for the digit sequences. Of most interest in these analyses is the effect of test appropriateness. Although there was only a marginal effect of this factor on typing latencies, it did have the expected strong effect on recognition responses. As revealed in Figure 22, subjects showed highest da scores for the sequences entered with the same key configuration at test as used at study, in accordance with the principle of transfer appropriate processing. Interestingly, when sequences were entered with a different response at study and at test, subjects' recognition memory was

no better than when they simply read the sequences at study. Entering the sequence at study only aided explicit recognition if the sequence was entered in the same way on the retention test. Therefore, this production effect can be seen as limited to the situation when the items were produced in the same way at study and at test.

Studies Demonstrating Considerable Forgetting

Memory for Numerical Calculations

Now we shall turn to our studies showing considerable forgetting over even relatively short retention intervals. In our work on memory for numerical calculations (Crutcher & Healy, 1989), we provided a more general test of the importance to memory of procedures or mental operations. This work followed from the phenomenon known as the generation effect. A growing number of experiments since the initial study by Slamecka and Graf (1978) have demonstrated a distinct retention advantage for material that is generated by an individual rather than simply read. If it is assumed that the generation effect is due to the activation in the subjects of auxiliary cognitive operations or mental procedures, then a task leading the subjects to perform such cognitive operations but not necessarily overt generation of an item may show equivalent retention to a generate task. Likewise, a task involving overt generation by the subjects but no auxiliary cognitive operations may not result in any better retention than a read task. In other words, according to this formulation it is not essential that the subjects generate or produce the stimulus, but rather it is essential that the subjects engage in the auxiliary cognitive operations or mental procedures linking the stimulus to other information stored in memory.

To test this cognitive operations hypothesis, we devised an experimental paradigm which allows for the orthogonal variation of stimulus presence (absent or present) and auxiliary cognitive operations (self or other). Four tasks are included, which we call the "read," "generate," "verify," and "calculate" tasks. Subjects in all four tasks are given single-digit multiplication problems. As

shown in Table 2, which presents a sample problem for each task, in the read task the answers are present in the problems and the multiplication operations are performed by another agent (the experimenter), whereas in the generate task the answers are absent and the multiplication operations are performed by the subjects themselves. The verify and calculate tasks are the ones crucial for testing the cognitive operations hypothesis. In the verify task the subjects are given a problem with its answer but are required to verify that the answer is correct. Contrastingly, in the calculate task, the subjects must provide the answers to the problems but they are told to use a calculator rather than perform the arithmetic themselves. After completing all the problems, subjects were asked to recall the answers to all the problems they had been shown. The cognitive operations hypothesis yields the prediction that retention on the verify and generate tasks would be superior to that on the read and calculate tasks, because in the former two tasks the multiplication operations are performed by the subjects themselves whereas in the latter two tasks the multiplication operations are performed by another agent (either the experimenter or a calculator). In contrast, no difference is expected between the generate and verify tasks or between the calculate and read tasks, because the difference between whether the answers are absent or present in the problems is not thought to be of much consequence. The results of primary interest are summarized in Table 3 in terms of proportions of correct responses on the free recall test. In accordance with our hypothesis, recall was greatly affected by whether or not the subjects performed the mental operations themselves but not by whether they were shown the answers with the problems.

In a follow-up experiment we aimed to assess the generalizability of these results. Our goal was to replicate and extend the findings from our first experiment along two dimensions. First, we sought to determine whether the same pattern of results would obtain for retention over considerably longer delays than were involved in the immediate testing situation of the first experiment.

Table 2

Illustration of Sample Problems for the Four Tasks in the Study of Memory for Numerical Calculations by Crutcher and Healy (1989)

Task	Subject Sees	Calculator Available	Subject Responds
Read	6 x 8 = 48	No	"6 x 8 = 48"
Generate	6 x 8 = ?	No	"6 x 8 = 48"
Calculate	6 x 8 = ?	Yes	"6 x 8 = 48"
Verify	6 x 8 = 48	No	"6 x 8 = 48, correct"

Table 3

Proportion of Correct Responses on the Free Recall Test as a Function of Cognitive Operations and Stimulus Presence in Experiment 1 of the Study of Memory for Numerical Calculations by Crutcher and Healy (1989)

	Cognitive Operations	
Stimulus Presence	Self	Other
	Verify	Read
Present	.68	.38
	Generate	Calculate
Absent	.68	.42

Second, we aimed to assess whether a recognition test procedure would lead to the same findings as the recall procedure used in the first experiment.

The method was similar to that in the first experiment except that subjects were tested either immediately, after a two-day delay, or after a seven-day delay. Right after the recall task subjects were given the recognition test. Subjects were shown pairs of multiplication products and for each pair they were to circle the one number in the pair that was an answer to one of the multiplication problems they were given during the study phase.

The results of the recall task are summarized in Table 4 in terms of proportions of correct recall responses for the four tasks in each of the three retention interval conditions. As in Experiment 1, recall levels for the generate and verify conditions were higher than those for the read and calculate conditions, and the same pattern of results was found for each of the three retention interval conditions although increased delay between study and test did depress performance levels considerably.

The results of the forced-choice recognition task are summarized in Table 5 in terms of proportions of correct recognition responses. Although performance levels for the recognition task were higher than for the recall task, the same pattern of results was found for recognition as for recall. Specifically, recognition levels were higher for shorter delays between study and test and, most crucially, were higher for the generate and verify conditions than for the read and calculate conditions, with essentially no differences between the generate and verify or between the read and calculate conditions.

It is important to note that although performance was influenced by the use of cognitive operations in this task, in no case was performance at the ceiling, and we did find substantial decreases in performance over retention intervals up to a week, so that permastore contributes little to these results.

Vocabulary Learning

Although we did not provide a direct test of our hypotheses in our studies

Table 4

Proportion of Correct Responses on the Free Recall Test as a Function of Cognitive Operations, Stimulus Presence, and Retention Interval in Experiment 2 of the Study of Memory for Numerical Calculations by Crutcher and Healy (1989)

Stimulus Presence and Retention Interval	Cognitive Operations: Self	Cognitive Operations: Other
Present	Verify	Read
Immediate	.59	.42
Two-day	.40	.24
Seven-day	.24	.10
Mean	.41	.25
Absent	Generate	Calculate
Immediate	.55	.34
Two-day	.49	.16
Seven-day	.40	.14
Mean	.48	.21

Table 5

Proportion of Correct Responses on the Recognition Test as a Function of Cognitive Operations, Stimulus Presence, and Retention Interval in Experiment 2 of the Study of Memory for Numerical Calculations by Crutcher and Healy (1989)

Stimulus Presence	Cognitive Operations	
and Retention Interval	Self	Other
Present	Verify	Read
Immediate	.82	.81
Two-day	.76	.61
Seven-day	.72	.52
Mean	.77	.65
Absent	Generate	Calculate
Immediate	.81	.65
Two-day	.75	.66
Seven-day	.69	.64
Mean	.75	.65

of other task domains, the results in these other domains are also consistent with our theoretical framework. For example, our work with vocabulary retention (Crutcher, 1989; Crutcher & Ericsson, 1988) provides important support for the hypothesis that the durability of memory depends critically on the extent to which the encoding generated during learning can be reinstated at test. This work is also of particular relevance to the conception of permastore, because it involves the same domain studied by Bahrick (1984). In these studies subjects were given the task of learning Spanish-English vocabulary items using the keyword method, which is a two-part mnemonic procedure which has been shown in the past to be effective in the teaching of foreign vocabulary. This method is illustrated in Figure 23. In this example, subjects are given the Spanish word (doronico), a mediating keyword (door) and the English translation (leopard). The subject's task is first to form a phonological (or orthographic) link between the Spanish word and the keyword and then to form an interactive image linking the keyword and the English word. In our initial experiments with this method, we tested the subjects on each of the two subtasks as well as the complete task. Specifically, in the phonological test we provided the Spanish word as a cue and required subjects to produce the keyword as a response; in the image test we provided the keyword as a cue and required subjects to produce the English word as a response; and in the full task the Spanish word was provided as a cue and subjects were required to produce the English word as a response. In our first experiment we tested two groups of subjects, both of whom acquired the vocabulary items followed by an immediate test. One group of subjects was then retested after a one-week retention interval, whereas the other group was retested after a one-month delay. The results for the two groups of subjects on the immediate and delayed tests are summarized in Figure 24 in terms of proportions of correct responses. Although there was little difference among the three tasks on the immediate test, the phonological task (in which subjects are given the Spanish word and respond with the keyword) was best retained after

the delay, especially for the one-month delay interval. This result suggests that it is the image component of the full task which is largely responsible for the forgetting observed when the keyword method is employed. We propose as an explanation for the difference between the phonological and image components that the procedures used in the phonological task to generate the keyword from the Spanish word are easily reinstated at the time of the test, whereas the procedures used in the image task to generate the English word from the keyword are not easily reinstated. In other words, subjects given the Spanish word can easily use inferencing to generate the keyword by thinking of words that sound like the Spanish word, because there are only a limited number of words that meet this relatively strict phonological constraint. On the other hand, subjects given the keyword cannot easily derive the English word by thinking of words that can be imaged together with the keyword, because there are a very large if not infinite number of different words that meet this much looser image constraint.

Two new experiments by Crutcher and Ericsson provide support for this explanation. In the first experiment we recalled 4 of the original 24 subjects from our earlier study, two initially tested after one week and two after one month. In the present experiment these subjects were tested after a retention interval of approximately one year. In this case, forgetting would be expected to be considerable, so that subjects would have to rely more on the inference or generating procedures than they would at the earlier tests. Indeed it was found, as shown in Figure 25, that performance of these subjects on the image and full tasks is very depressed, whereas performance on the phonological task, though clearly lower than that after the initial delayed tests, is still extremely high. Note, however, that even for the image and full tasks, there is some retention after a year. This observation is consistent with Bahrick's (1984) findings concerning permastore, especially given the relatively limited amount of practice by our subjects.

In the second experiment, a new group of subjects was asked to generate keywords given the same Spanish words as used previously. The subjects were given 40 seconds for each item to produce as many keywords as possible. The subjects were told to generate words that were phonologically or orthographically similar to the Spanish words. Both the proportion of trials on which the original keywords were produced by the subjects as their first response and the proportion of trials on which the original keywords were produced as any response within the 40 seconds were computed and are shown in Figure 26. Figure 26 includes a conservative estimate of subjects' retention of the keywords after one year. This estimate is based on recall accuracy for the first trials of the retention test before the subjects were exposed to the keywords as part of any other test stimuli. On only 15% of the trials did subjects produce an incorrect response, which implies substantial memory of the keywords even after corrections for guessing. These results suggest that the subjects primarily relied on a generate-and-test method to retrieve the presented keywords.

Components of Memory for Course Schedules

Our last task domain involves the recall by undergraduate students of information about their course schedules. In this study (Wittman, 1989), we examined the retention of four different types of course schedule information -- the name of the course ("what"), the instructor's name ("who"), the location of the class building on a campus map ("where"), and the class start time ("when"). In our first experiment we used both a cross-sectional and longitudinal design to assess the relative forgetting rates of these four different types of information. We tested three groups of subjects on three different occasions, with the three tests occurring on consecutive semesters. Subjects were initially tested after completing at least two years of course work at the University of Colorado. During each of the three tests they were questioned about their course schedule from a previous semester, with a different semester

tested on the three occasions, so that on average there was a six-month retention interval separating the three tests. The testing made use of a cued recall procedure. Subjects were probed about three courses taken during the test semester. Across subjects each type of information was used equally often as a cue to recall the other three types of information. The results are summarized in Figure 27 in terms of mean percentages of correct responses for the four different types of information at the three testings. There was considerable forgetting of this course schedule information. This finding is evident in two different ways. First, the overall levels of performance are quite low. For example, at the second and third testings subjects recall less than half of the time the course title, the instructor's name, and the class time. Second, there is a significant overall decrease in performance on the second and third tests relative to the first test. Hence, it is clear that the course schedule information, although learned naturally and with considerable reexposure during the semester in question, is not well retained, and certainly does not rely on permastore to any great extent. The most interesting observation concerns the differences among the four types of information. On all three tests, subjects' performance was much better on the spatial, or where, information than on the other three types (which is consistent with the finding that spatial information is retained better than temporal information in studies of short-term memory; see, e.g., Healy, 1975).

In accordance with the theoretical framework outlined earlier, we propose as an explanation for the superiority of spatial location recall in this case that subjects learned this information by using procedures which were repeated throughout the semester. Specifically, subjects walked through the campus to the classroom each time the class was held. A similar type of procedural learning was not as readily available for the course title, instructor's name, or class start time.

In order to provide an initial test of this hypothesis, we conducted a

follow-up experiment in which subjects had to learn course schedules in the laboratory. We compared two test conditions, the map test and the class-listing test, which differed in the amount of procedural memory required. During the study phase of the experiment subjects were provided with both a map of the campus and a course schedule in a format similar to that naturally provided to students at the university. This schedule included the four types of information studied in the first experiment along with some ancillary information such as the classroom number. Subjects were given nine training trials followed by a pair of retention tests one week later and then another pair of tests after approximately five more weeks. In both tests subjects were required to provide from memory the course title and the instructor's name. The tests differed in the type of temporal and spatial information required. In the map test the subjects provided the order of their classes during the school week and the location of each class on the campus map. In contrast, in the class-listing test, subjects provided the start time of each course and the building name where the class was held. The map test was meant to mimic as closely as we could with paper and pencil the procedures naturally used to retrieve course locations, whereas the class-listing test was meant to remove any procedural component from the recall of course locations.

The results are summarized in Figure 28 in terms of percentages of correct responses on our two tests as a function of retention interval (one week for the first test and six weeks for the retest) and information type. The comparison of the two retention intervals makes it clear that, as in the natural situation, subjects showed forgetting of the course schedule information overall. Of most interest is the observation that the superiority of spatial information occurred only in the map test, which involved procedural memory. In fact, there was almost no sign of forgetting the spatial information on the map test over the six-week retention interval. In the class-listing test, spatial information showed no superiority in retention at test and showed significant loss at

retest. Thus, we have initial support for our hypothesis that the superiority of spatial memory is due to the fact that procedures are used to learn that information and these procedures can be readily reinstated at test.

Final Comments

Although the theoretical framework we have proposed, which is centered around the notion of procedural reinstatement, is able to throw considerable light on the long-term maintenance of knowledge and skills, it can by no means account for all important retention phenomena. Other theoretical constructs are needed to provide a more complete account of long-term retention, and the constructs we have outlined need to be fleshed out in greater detail and require more rigorous experimental tests that also determine their generalizability to other tasks and situations. Nevertheless, we have been impressed with the wide variety of memory studies that already fit into this framework and the remarkably large range of forgetting rates found in these studies. Recall that information about course schedules was quickly forgotten, so that many of our subjects could not report, even in approximate form, the name of the instructor, the course title, or the class meeting time for a course they had taken previously, even though that information was learned in a natural setting, had importance to the individuals, and was presumably given considerable rehearsal during a semester-long interval. In contrast, for example, our two subjects who were given extensive training in the laboratory on a target-detection skill were able to maintain their performance level without any noticeable decrease at all over a period as long as 15 months. If procedural reinstatement can indeed help to explain these large differences in forgetting rates, the practical significance of this finding is substantial. The implication is clear that if we wish to retain knowledge or skills over a long delay interval, it is crucial that we make sure that the procedures we use when learning the information are reinstated at the time we need to recall the information.

References

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. Journal of Experimental Psychology: General, 113, 1-29.

Crutcher, R. J. (1989). The role of mediation in knowledge acquisition and retention: Learning foreign vocabulary using the keyword method. Unpublished masters thesis, University of Colorado, Boulder, CO.

Crutcher, R. J., & Ericsson, K. A. (1988, April). A componential analysis of the keyword method. Paper presented at the American Educational Research Association Convention, New Orleans, Louisiana.

Crutcher, R. J., & Healy, A. F. (1989). Cognitive operations and the generation effect. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 669-675.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. American Scientist, 70, 607-615.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), Coding processes in human memory (pp. 161-190). Washington, DC: Winston.

Fendrich, D. W., Healy, A. F., & Bourne, L. E. (1988, November). Long-term retention of procedural and episodic memory for digits. Paper presented at the 29th Annual Meeting of the Psychonomic Society, Chicago, Illinois.

Fendrich, D. W., Healy, A. F., & Bourne, L. E. (1989, November). Training and retention of simple mental multiplication. Poster presented at the 30th Annual Meeting of the Psychonomic Society, Atlanta, Georgia.

Gesi, A. T., Fendrich, D. W., Healy, A. F., & Bourne, L. E. (1989, April). Episodic and procedural memory for digit sequences. Paper presented at the Joint Annual Convention of the Western Psychological Association and the

Rocky Mountain Psychological Association, Reno, Nevada.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 501-518.

Healy, A. F. (1975). Coding of temporal-spatial patterns in short-term memory. Journal of Verbal Learning and Verbal Behavior, 14, 481-495.

Healy, A. F., Fendrich, D. W., & Proctor, J. D. (1990). Acquisition and retention of a letter-detection skill. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 270-281.

Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. Journal of Verbal Learning and Verbal Behavior, 23, 425-449.

Naylor, J. C., & Briggs, G. E. (1961). Long-term retention of learned skill: A review of the literature (ASD-TR-61-390). Wright-Patterson AFB, OH: Advanced Systems Division.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16, 519-533.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84, 127-190.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory, 4, 592-604.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

Wittman, W. T. (1989). A long-term retention advantage for spatial information learned naturally and in the laboratory. Unpublished doctoral dissertation, University of Colorado, Boulder, CO.

Figure Captions

Figure 1. Sample displays of each frame size in the target detection study by Healy et al. (1990).

Figure 2. Proportion of hits as a function of day of training and frame size in Experiment 1 of the target detection study by Healy et al. (1990), with data from the limited training group in the top panel and data from the extensive training group in the bottom panel.

Figure 3. Average response latency as a function of day of training and frame size in Experiment 1 of the target detection study by Healy et al. (1990), with data from the limited training group in the top panel and data from the extensive training group in the bottom panel.

Figure 4. Retention data from the limited and extensive training groups in Experiment 1 of the target detection study by Healy et al. (1990) compared to data from the control group, with proportion of hits in the top panel and avergage response latency in the bottom panel.

Figure 5. Proportion of hits as a function of day of training and frame size in Experiment 2 of the target detection study by Healy et al. (1990), with data from subject AG in the top panel and data from subject DS in the bottom panel. (R1 stands for the first retention test, and R2 for the second retention test.)

Figure 6. Average response latency as a function of day of training and frame size in Experiment 2 of the target detection study by Healy et al. (1990), with data from subject AG in the top panel and data from subject DS in the bottom panel. (R1 stands for the first retention test, and R2 for the second retention test.)

Figure 7. Average log response latency as a function of multiplication column and training session number for the acquisition trials in Experiment 1 of the mental multiplication study by Fendrich et al. (1989), with data from subject MM in the top panel and data from subject SM in the bottom panel.

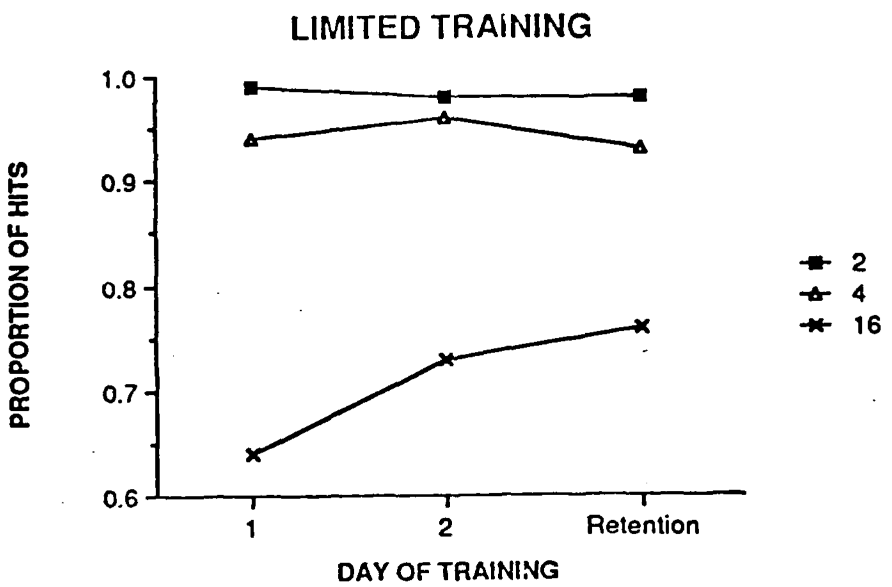
Figure 8. Average log response latency as a function of multiplication column

and session number for the last acquisition and retention trials in Experiment 1 of the mental multiplication study by Fendrich et al. (1989), with data from subject MM in the top panel and data from subject SM in the bottom panel.

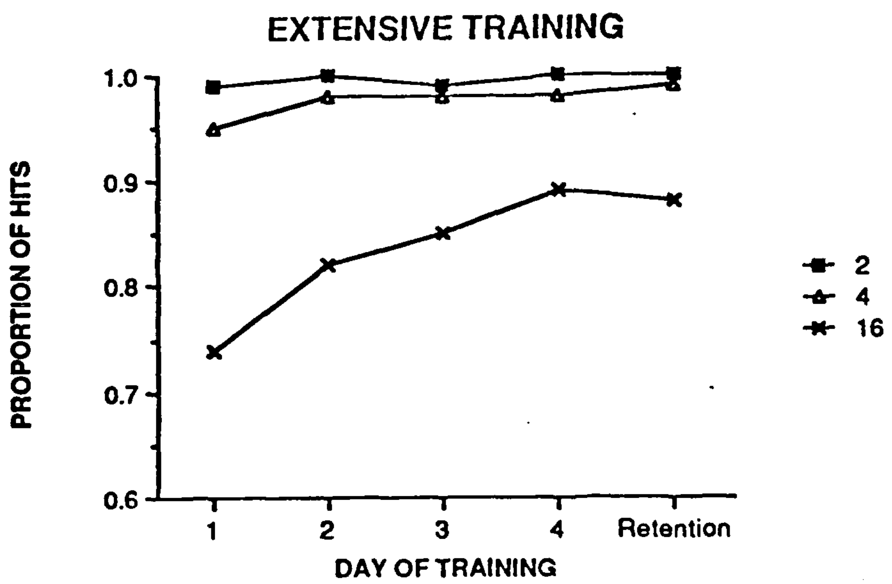
information type in Experiment 2 of the study of components of memory for course schedules by Wittman (1989), with results of the map test in the top panel and results of the class listing test in the bottom panel.

FRAME SIZE

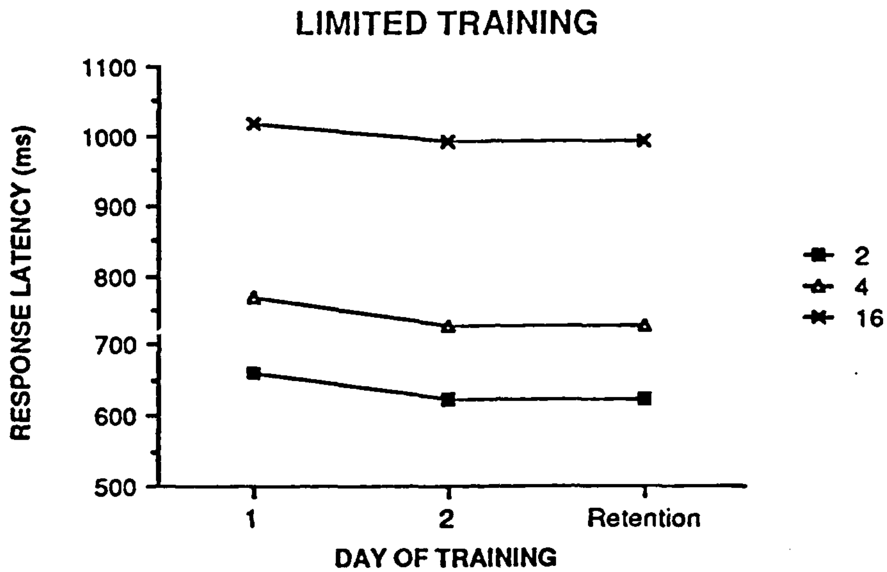
2	## ####### ######HI#
4	##O ## H####I##M###
16	WYSEYIG PEO PCNUHE

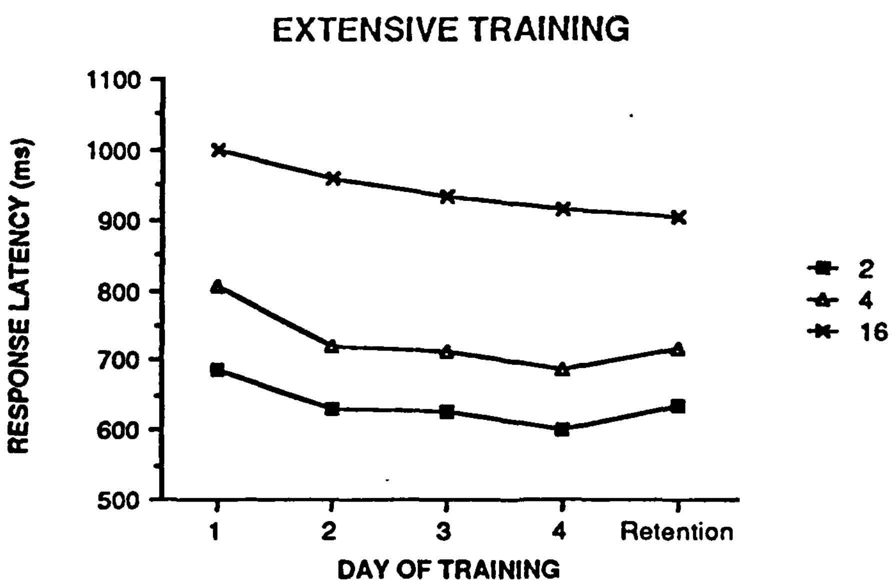


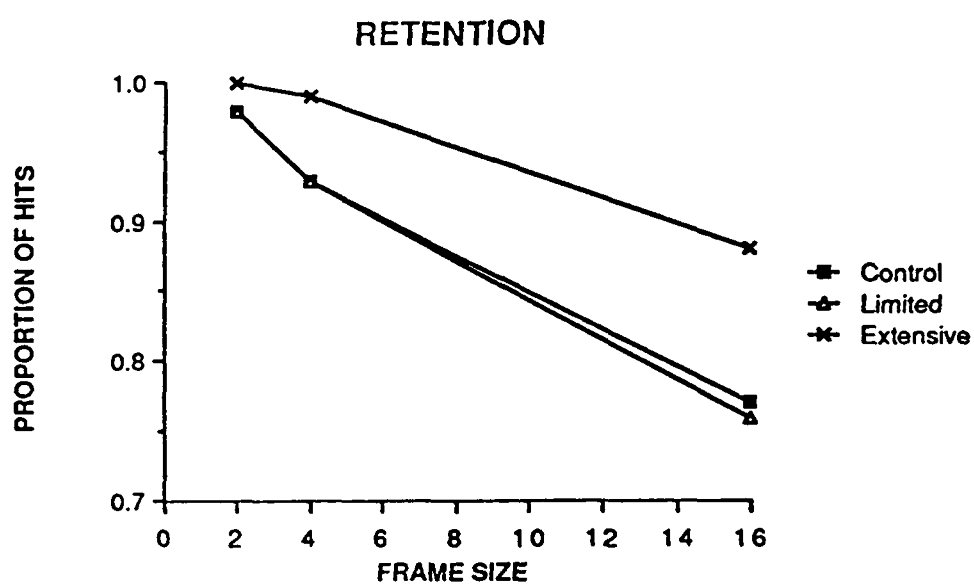
LIMITED TRAINING
PROPORTION OF HITS
1.0
0.9
0.8
0.7
0.6
1
2
Retention
DAY OF TRAINING
2
4
16

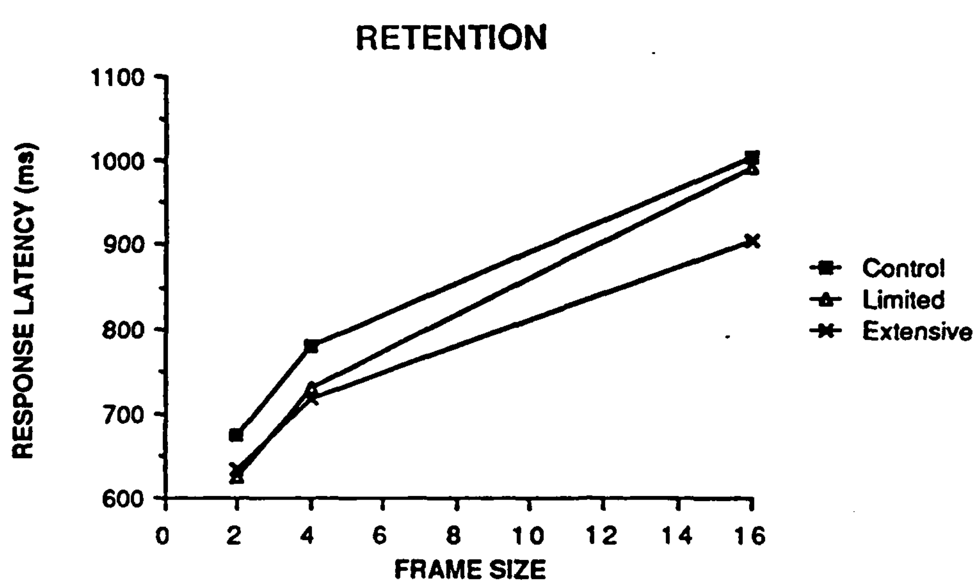


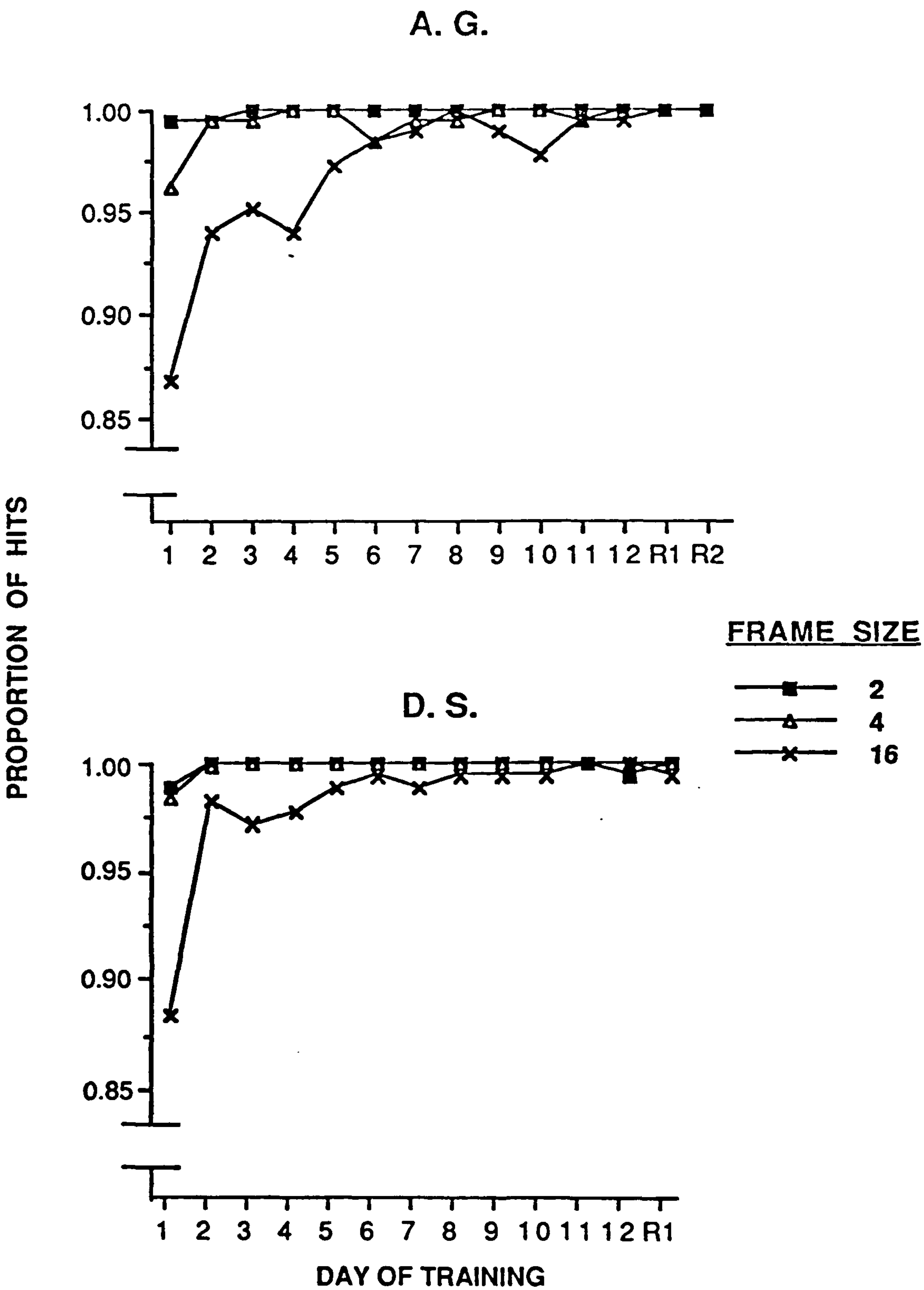
EXTENSIVE TRAINING
PROPORTION OF HITS
1.0
0.9
0.8
0.7
0.6
1
2
3
4
Retention
DAY OF TRAINING
2
4
16

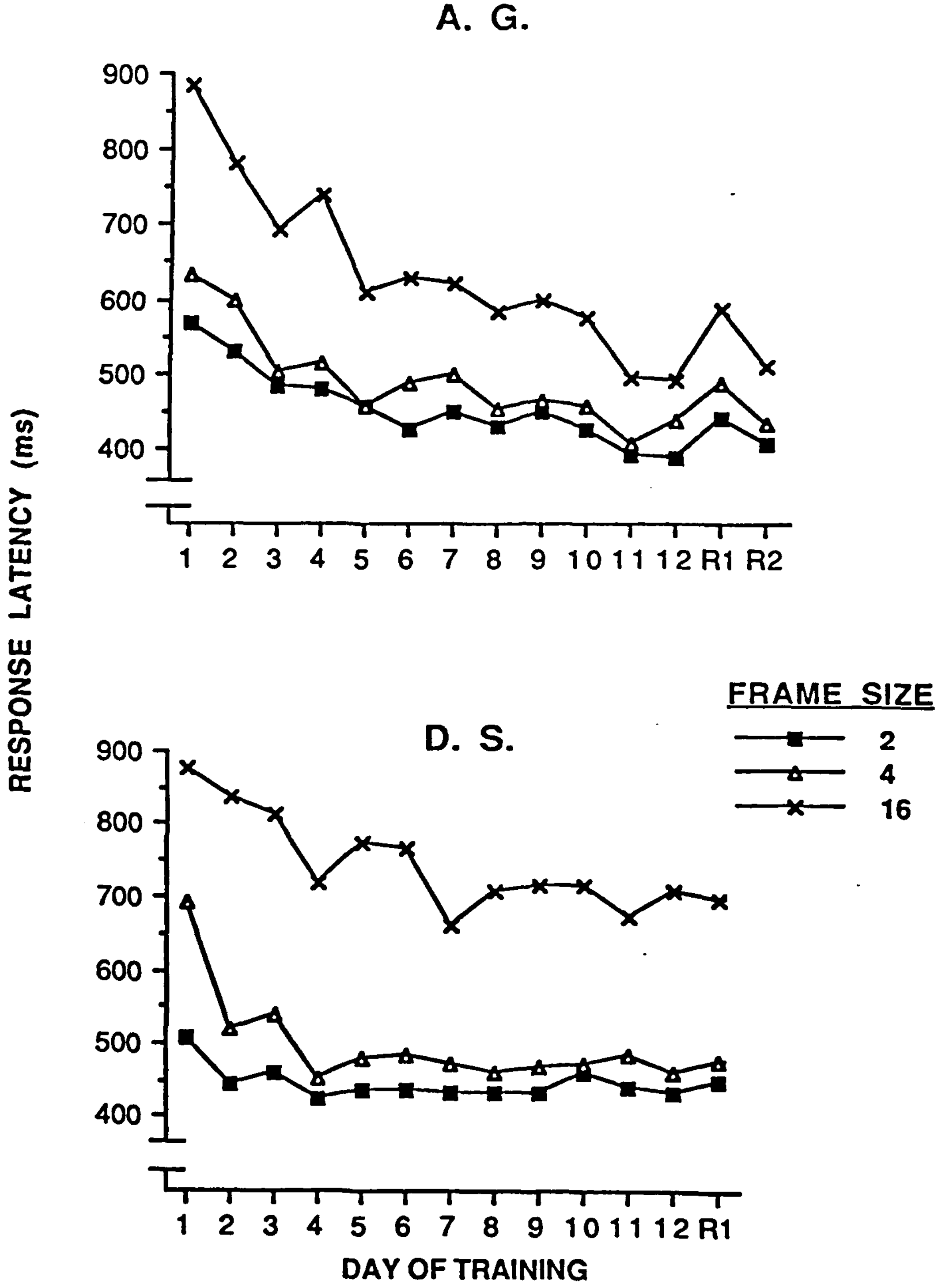

LIMITED TRAINING
1100
1000
900
800
700
600
500
RESPONSE LATENCY (ms)
1
2
Retention
DAY OF TRAINING
2
4
16

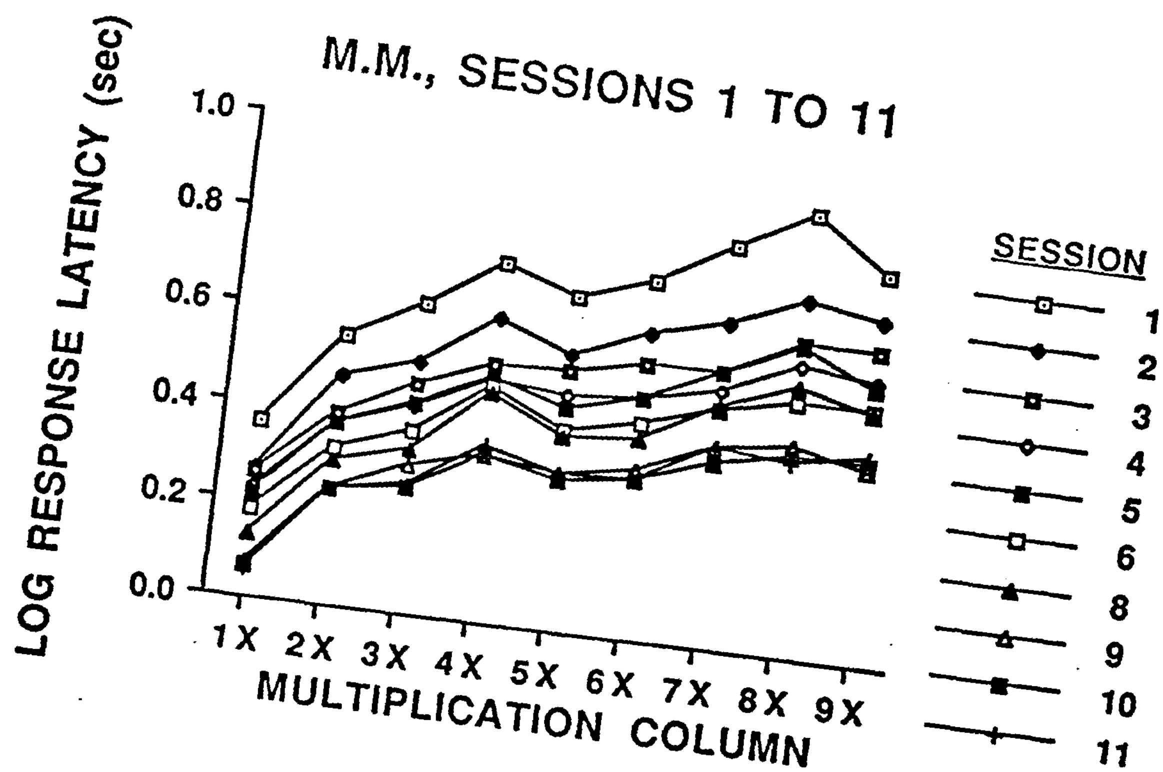

EXTENSIVE TRAINING
1100
1000
900
800
700
600
500
RESPONSE LATENCY (ms)
1
2
3
4
Retention
DAY OF TRAINING
2
4
16


RETENTION
PROPORTION OF HITS
1.0
0.9
0.8
0.7
0
2
4
6
8
10
12
14
16
FRAME SIZE
Control
Limited
Extensive

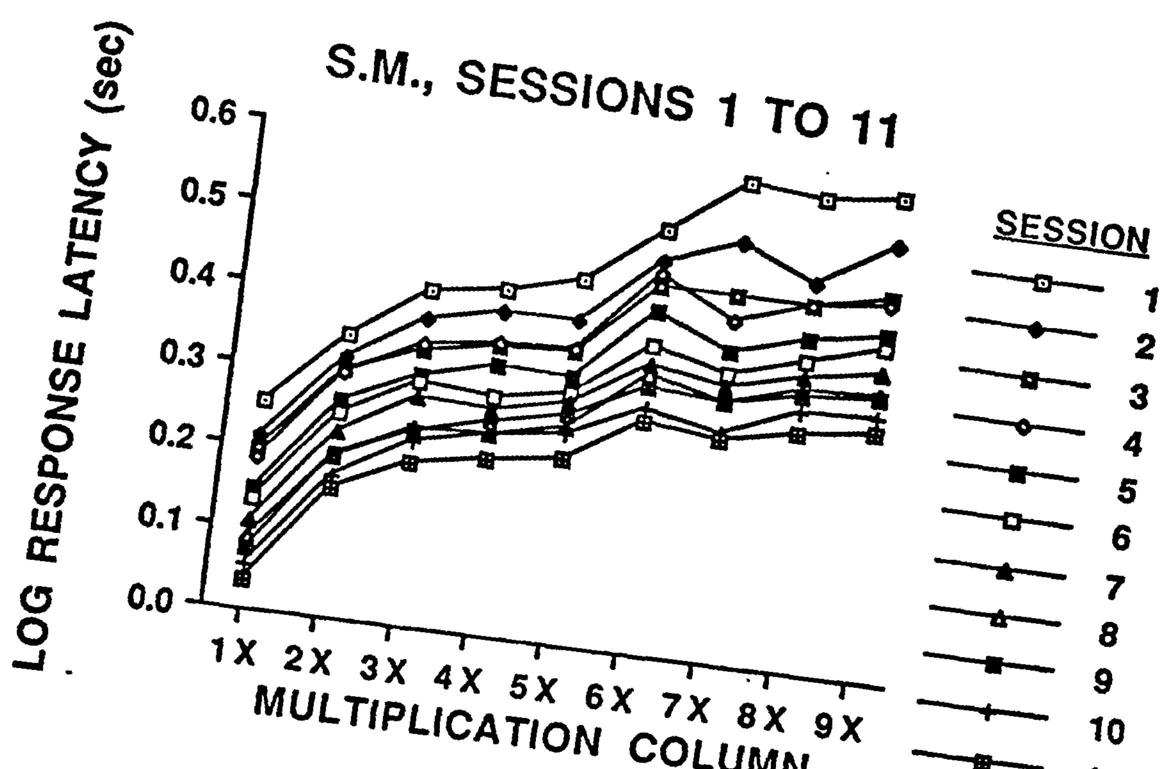

RETENTION
RESPONSE LATENCY (ms)
1100
1000
900
800
700
600
0
2
4
6
8
10
12
14
16
FRAME SIZE
Control
Limited
Extensive


A. G.
1.00
0.95
0.90
0.85
1 2 3 4 5 6 7 8 9 10 11 12 R1 R2
PROPORTION OF HITS
FRAME SIZE
2
4
16
D. S.
1.00
0.95
0.90
0.85
1 2 3 4 5 6 7 8 9 10 11 12 R1
DAY OF TRAINING

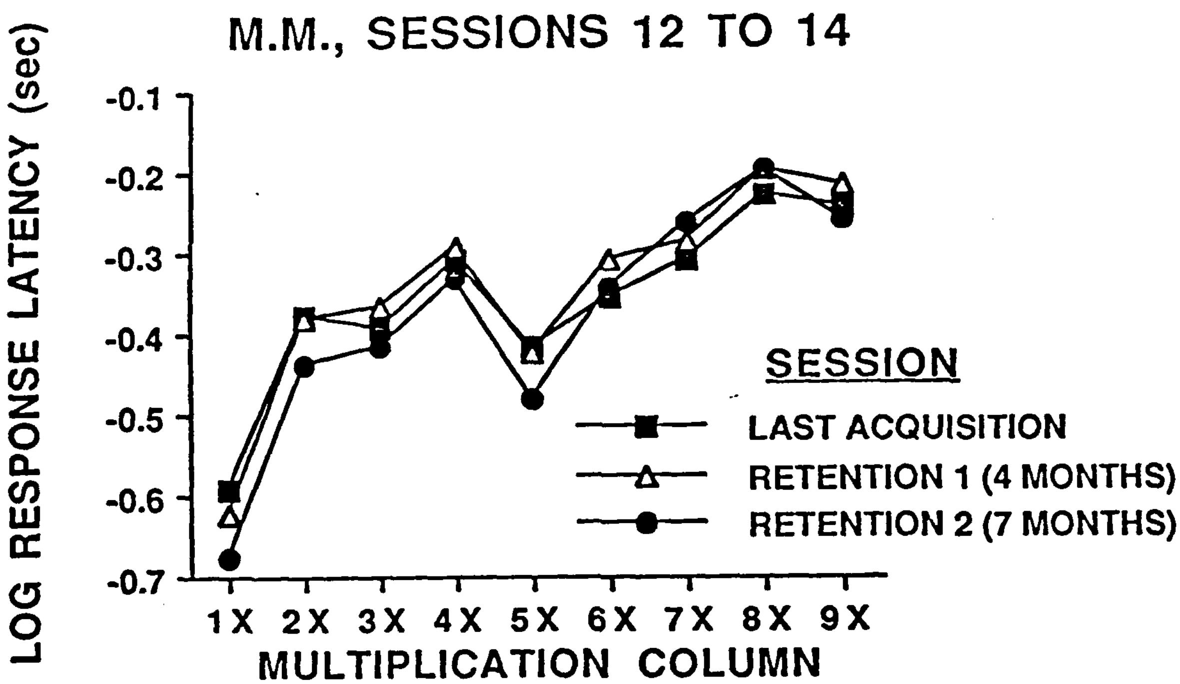

A. G.
D. S.
RESPONSE LATENCY (ms)
900
800
700
600
500
400
1 2 3 4 5 6 7 8 9 10 11 12 R1 R2
1 2 3 4 5 6 7 8 9 10 11 12 R1
DAY OF TRAINING
FRAME SIZE
2
4
16

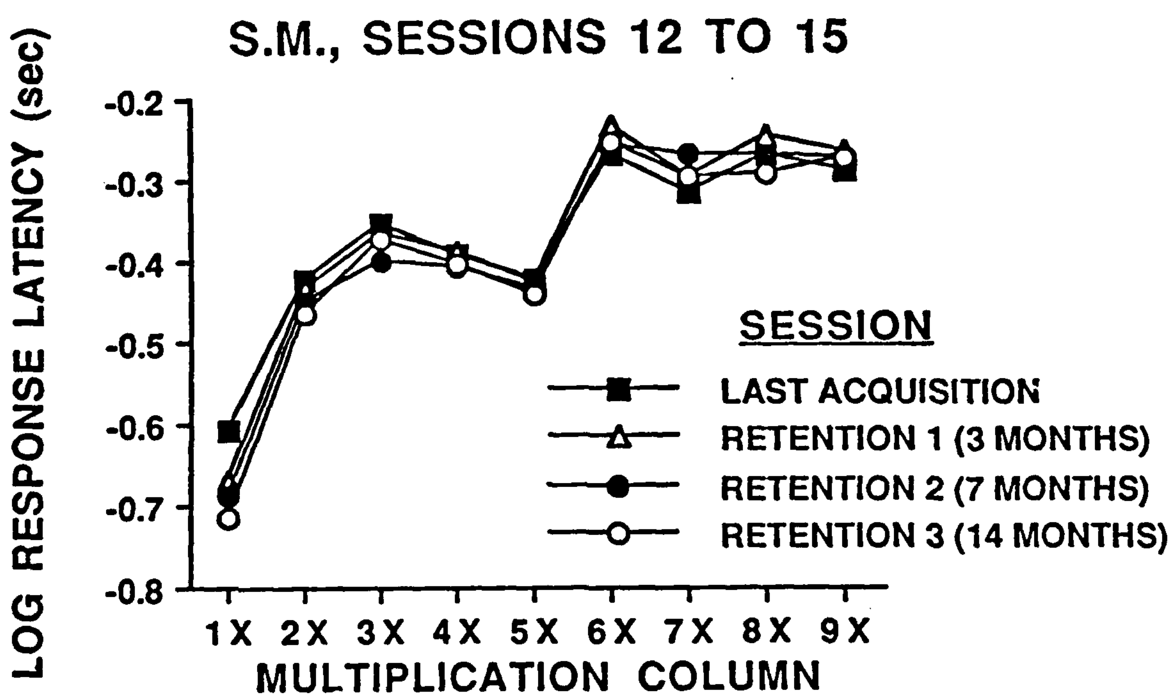


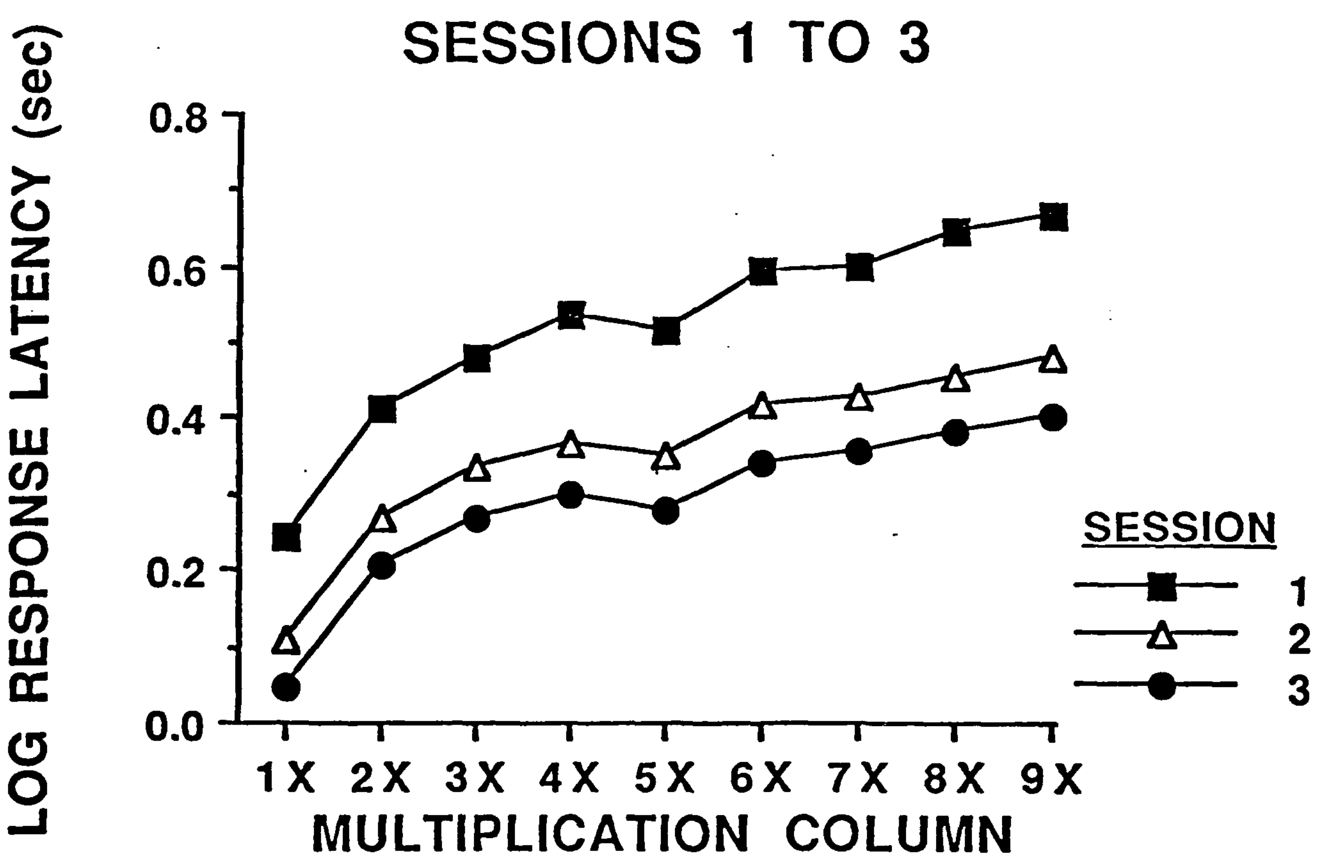
M.M., SESSIONS 1 TO 11
LOG RESPONSE LATENCY (sec)
1.0
0.8
0.6
0.4
0.2
0.0
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
1
2
3
4
5
6
8
9
10
11



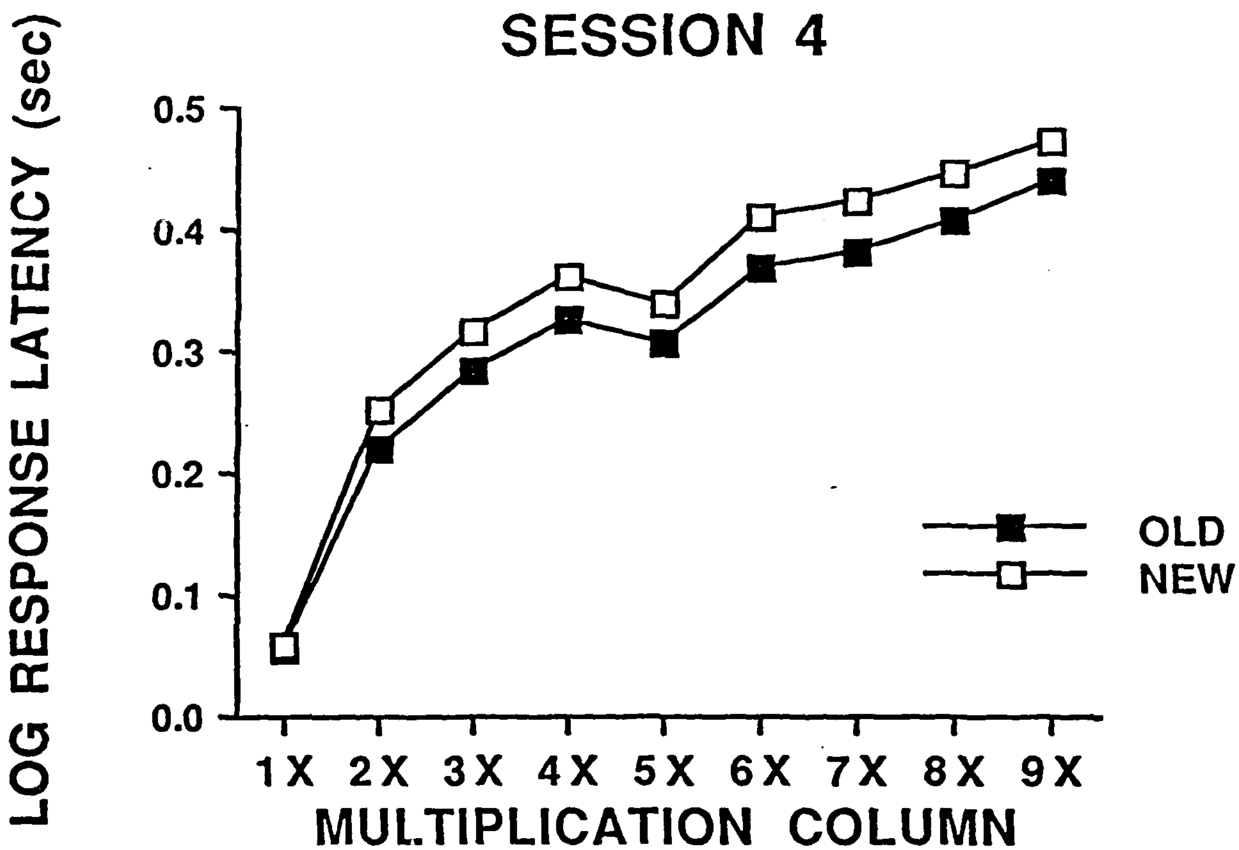
S.M., SESSIONS 1 TO 11
LOG RESPONSE LATENCY (sec)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
1
2
3
4
5
6
7
8
9
10
11

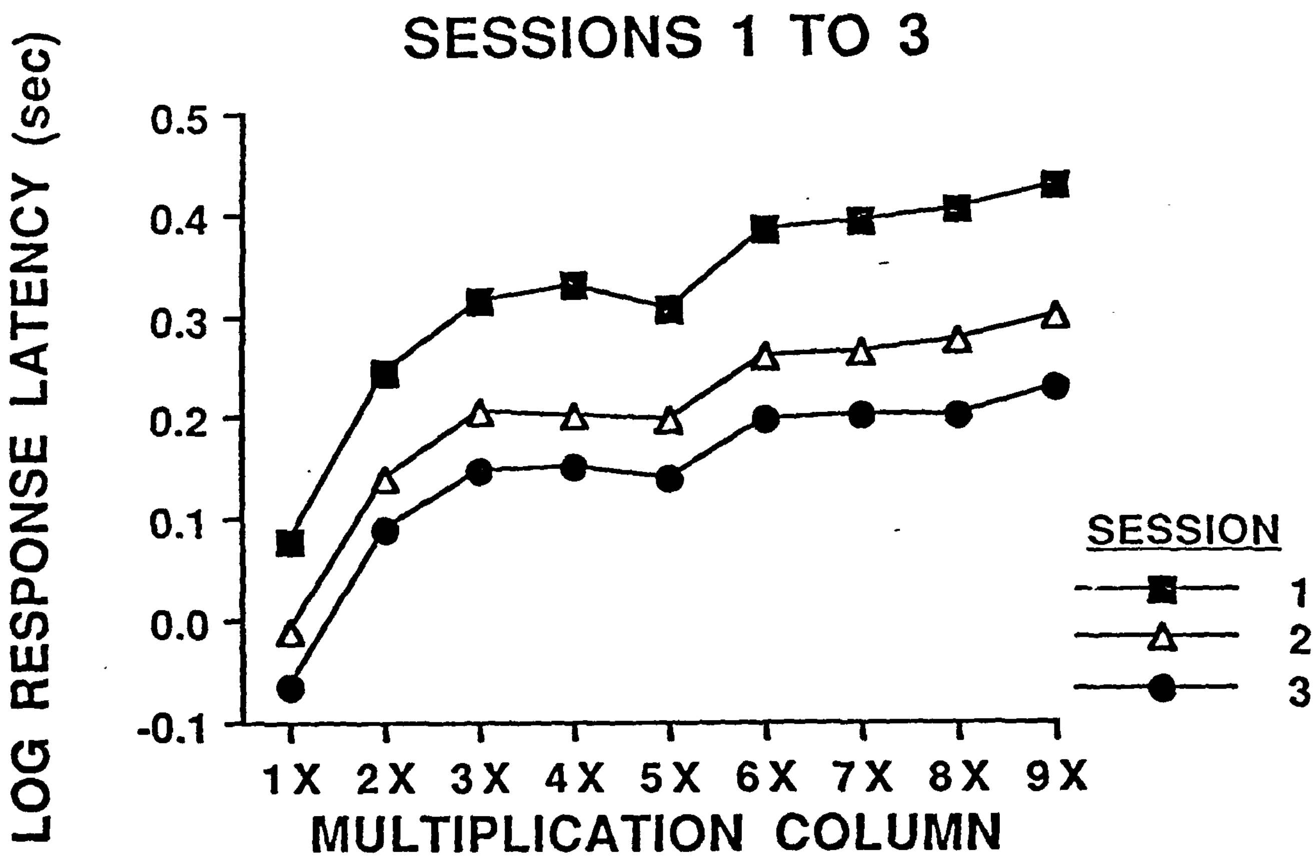

M.M., SESSIONS 12 TO 14
LOG RESPONSE LATENCY (sec)
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
LAST ACQUISITION
RETENTION 1 (4 MONTHS)
RETENTION 2 (7 MONTHS)

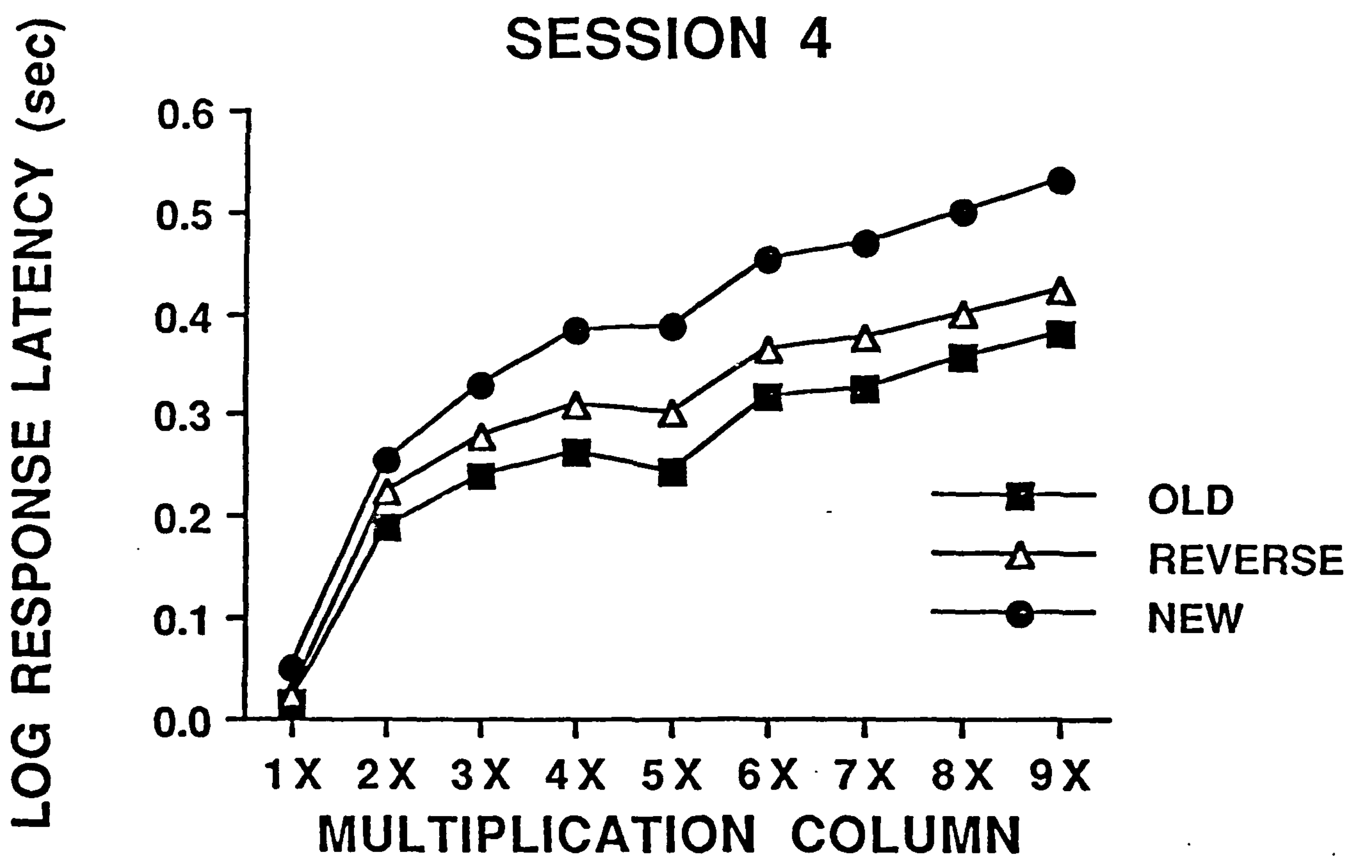

S.M., SESSIONS 12 TO 15
LOG RESPONSE LATENCY (sec)
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
LAST ACQUISITION
RETENTION 1 (3 MONTHS)
RETENTION 2 (7 MONTHS)
RETENTION 3 (14 MONTHS)



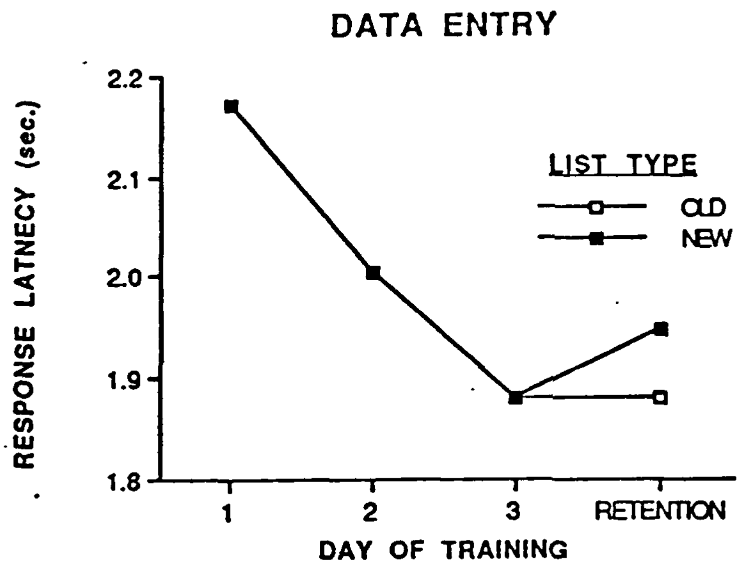
SESSIONS 1 TO 3
LOG RESPONSE LATENCY (sec)
0.8
0.6
0.4
0.2
0.0
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
1
2
3

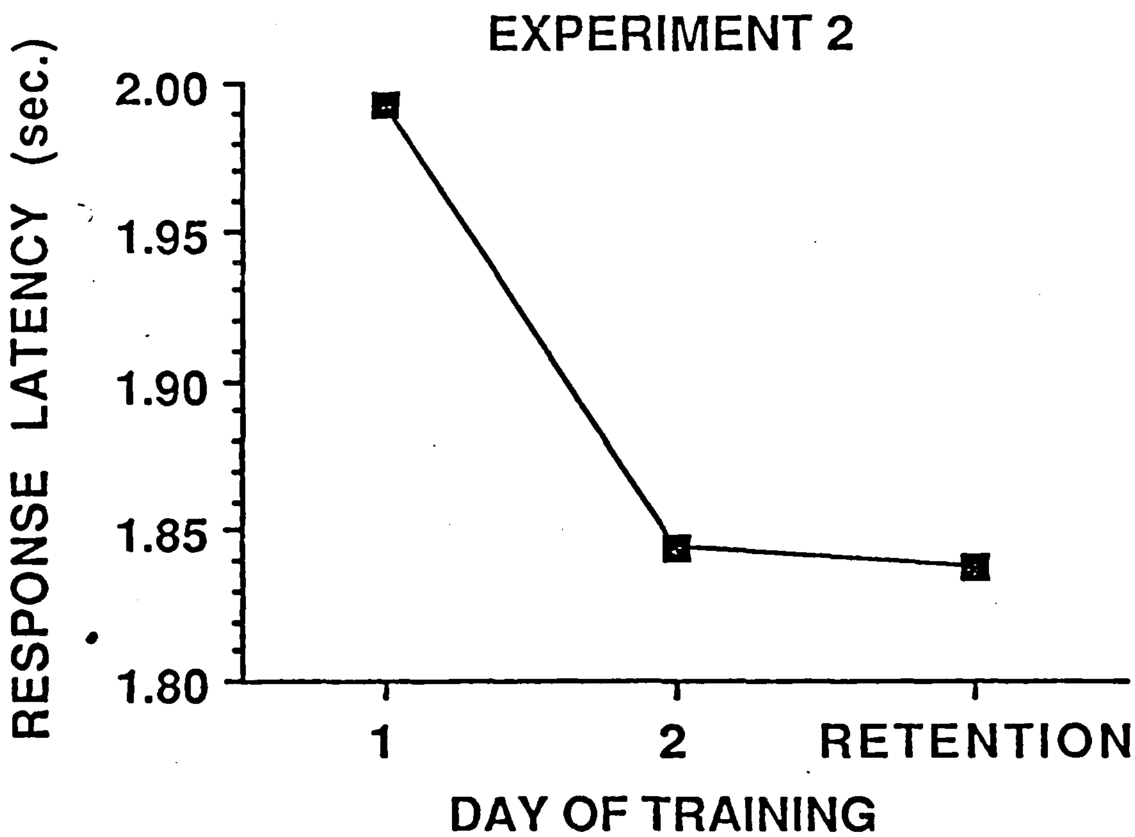

SESSION 4
LOG RESPONSE LATENCY (sec)
0.5
0.4
0.3
0.2
0.1
0.0
1X
2X
3X
4X
5X
6X
7X
8X
9X
MULTIPLICATION COLUMN
OLD
NEW

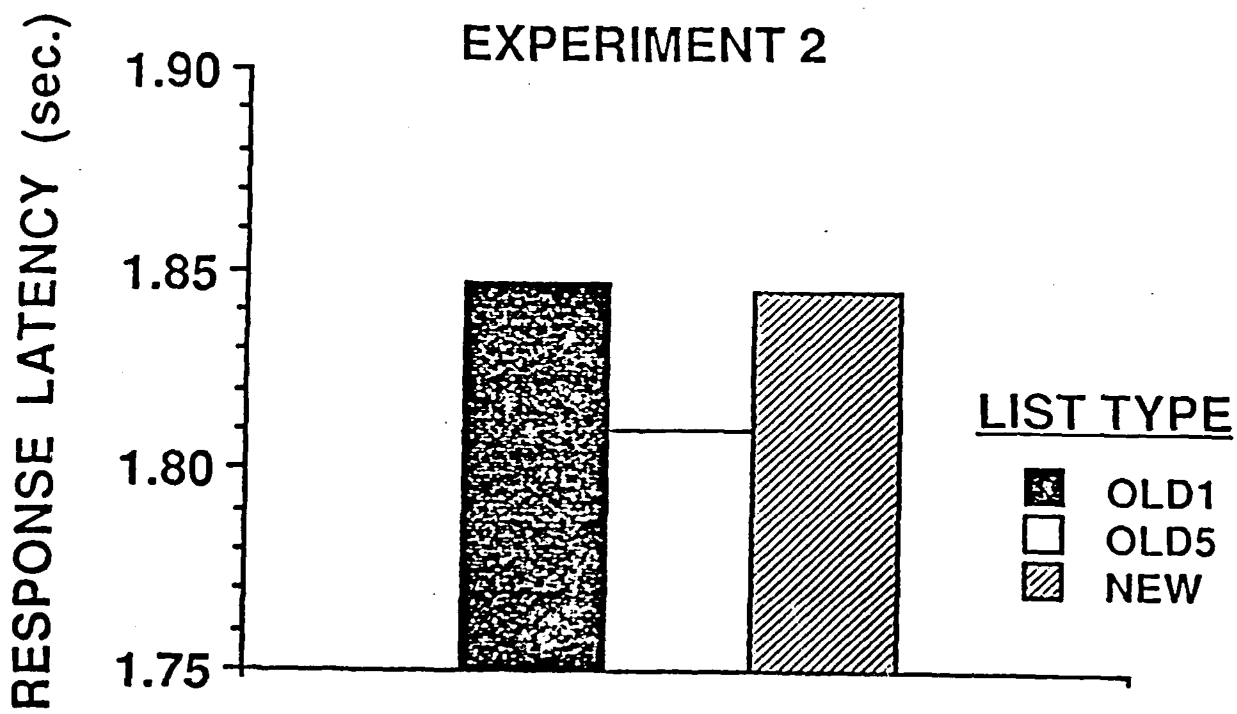

SESSIONS 1 TO 3
LOG RESPONSE LATENCY (sec)
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
SESSION
1
2
3

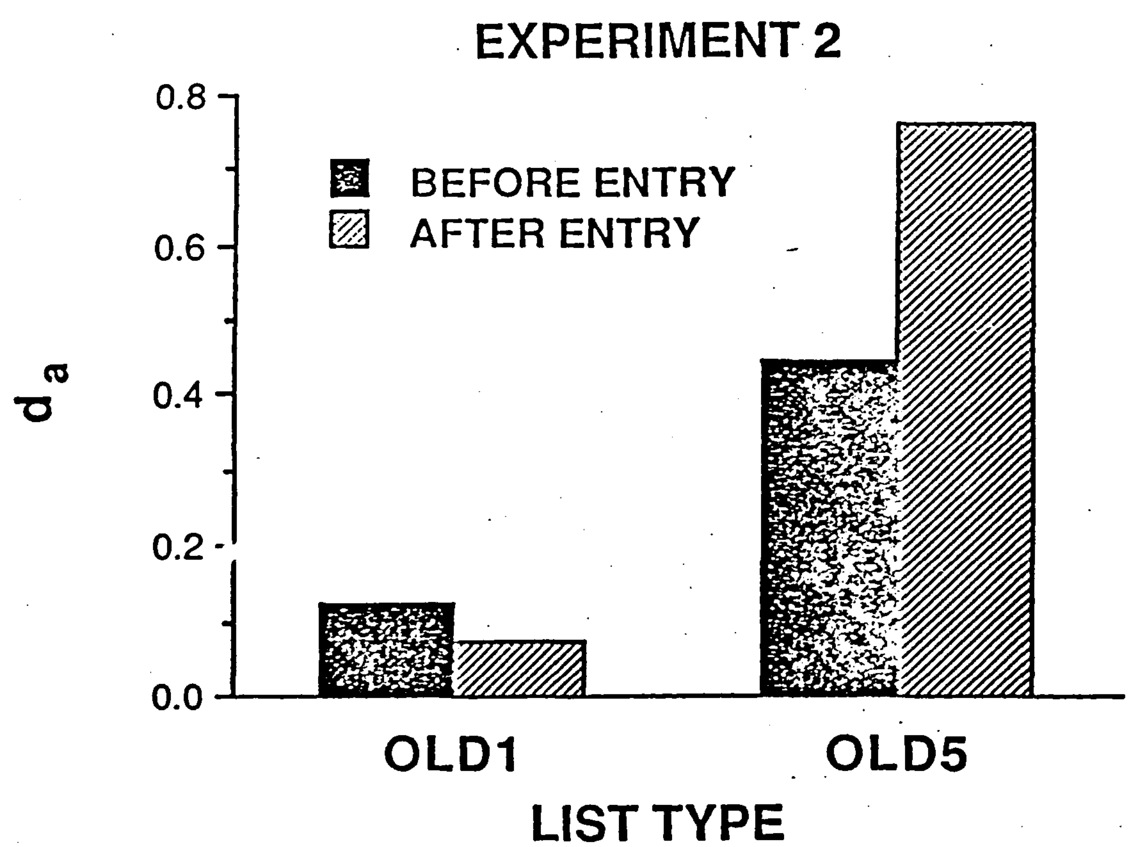


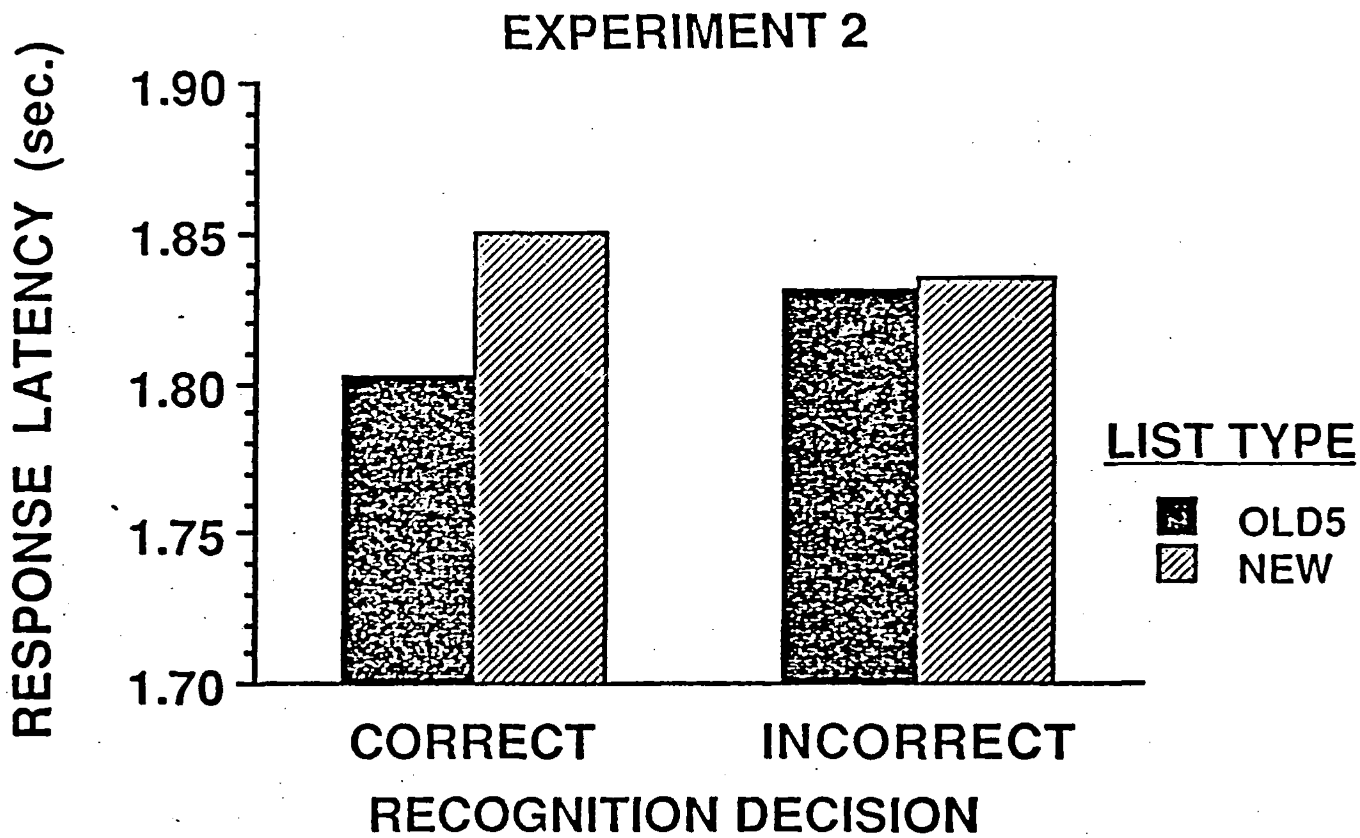
SESSION 4
LOG RESPONSE LATENCY (sec)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1X 2X 3X 4X 5X 6X 7X 8X 9X
MULTIPLICATION COLUMN
OLD
REVERSE
NEW


DATA ENTRY
RESPONSE LATNECY (sec.)
2.2
2.1
2.0
1.9
1.8
LIST TYPE
OLD
NEW
1
2
3
RETENTION
DAY OF TRAINING

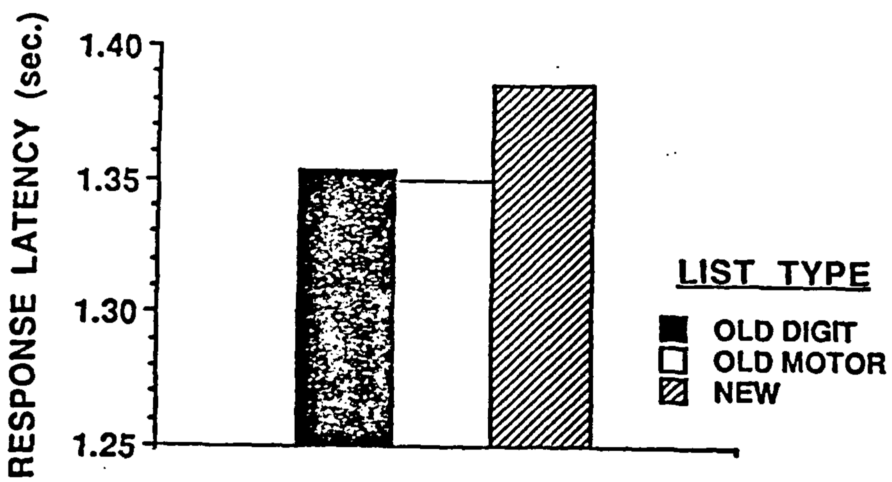

EXPERIMENT 2
RESPONSE LATENCY (sec.)
2.00
1.95
1.90
1.85
1.80
1
2
RETENTION
DAY OF TRAINING

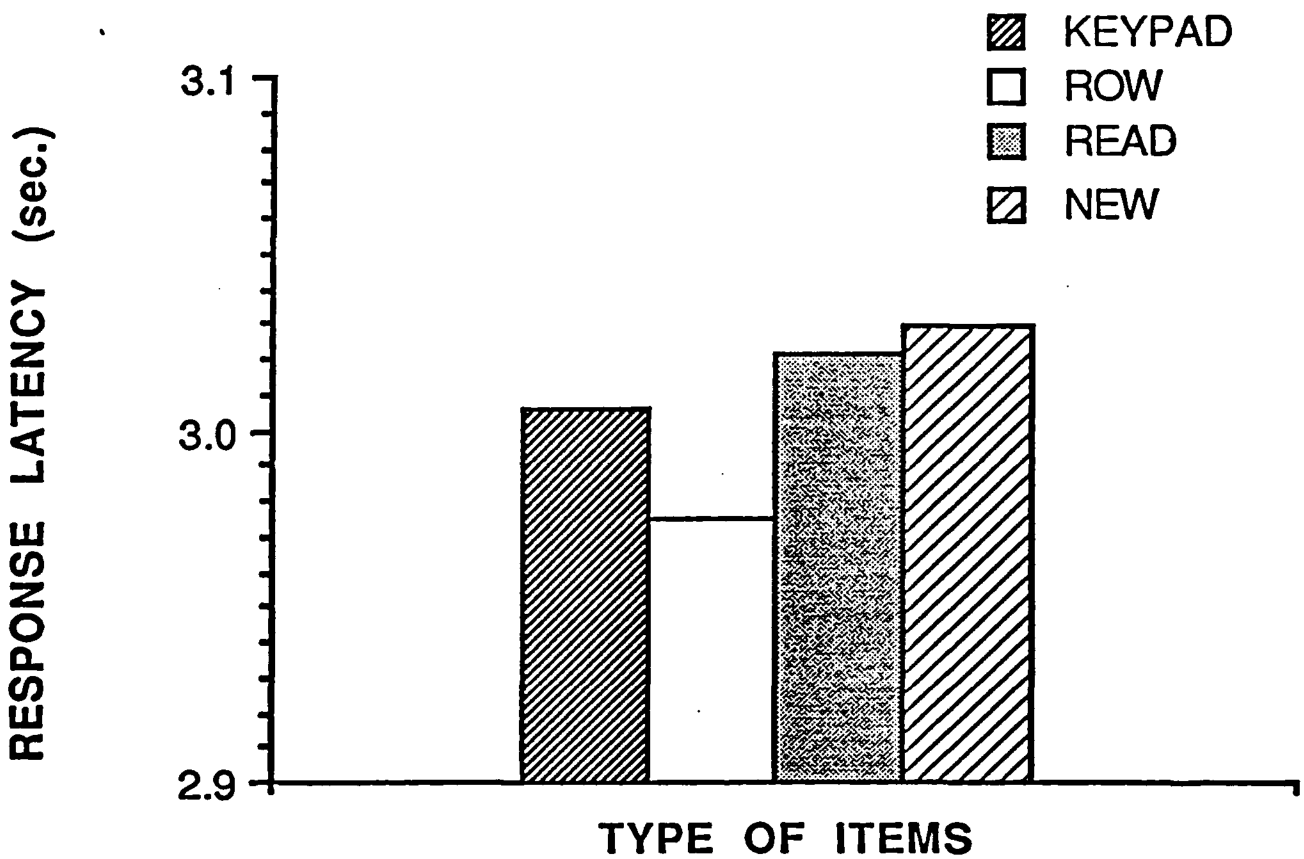

EXPERIMENT 2
RESPONSE LATENCY (sec.)
1.90
1.85
1.80
1.75
LIST TYPE
OLD1
OLD5
NEW

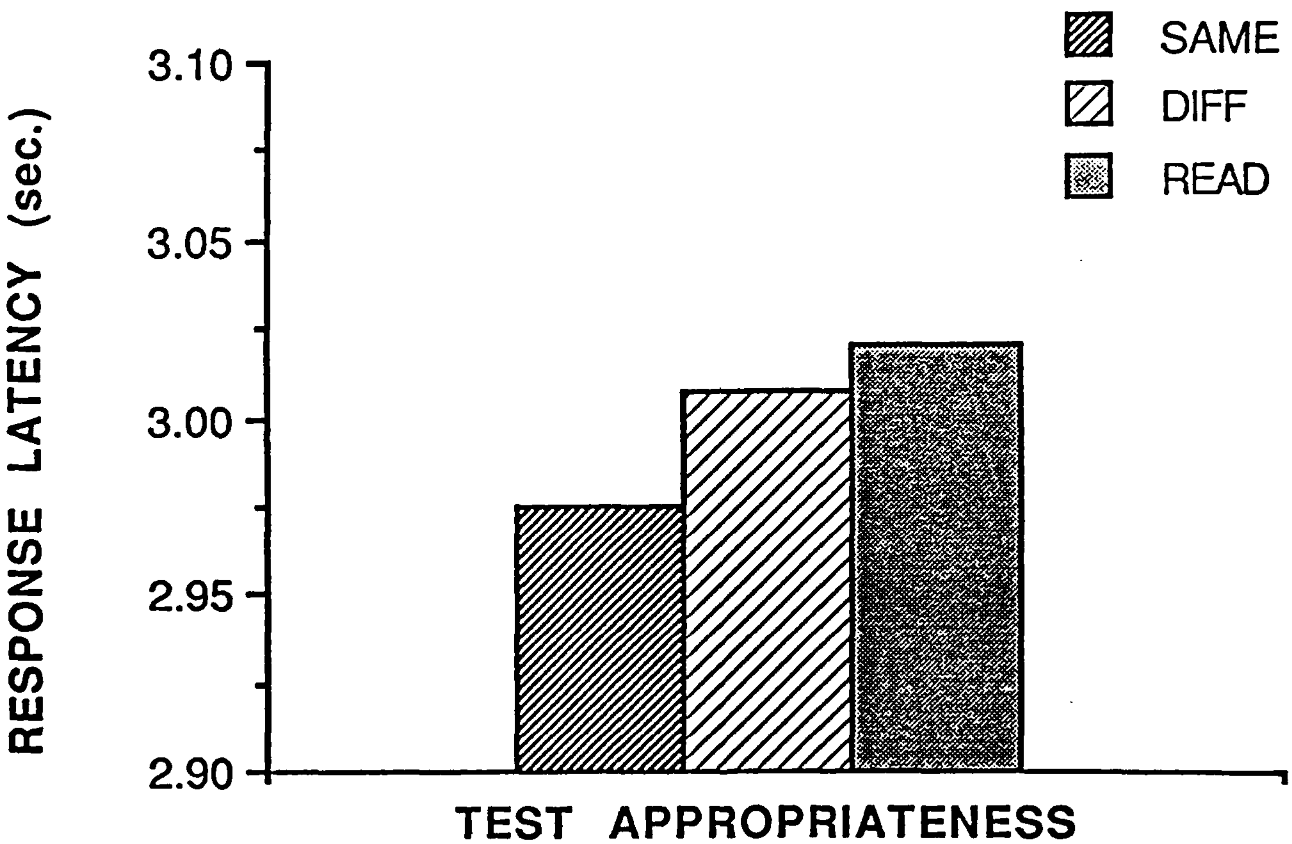

EXPERIMENT 2
0.8
0.6
0.4
0.2
0.0
d a
BEFORE ENTRY
AFTER ENTRY
OLD1
OLD5
LIST TYPE

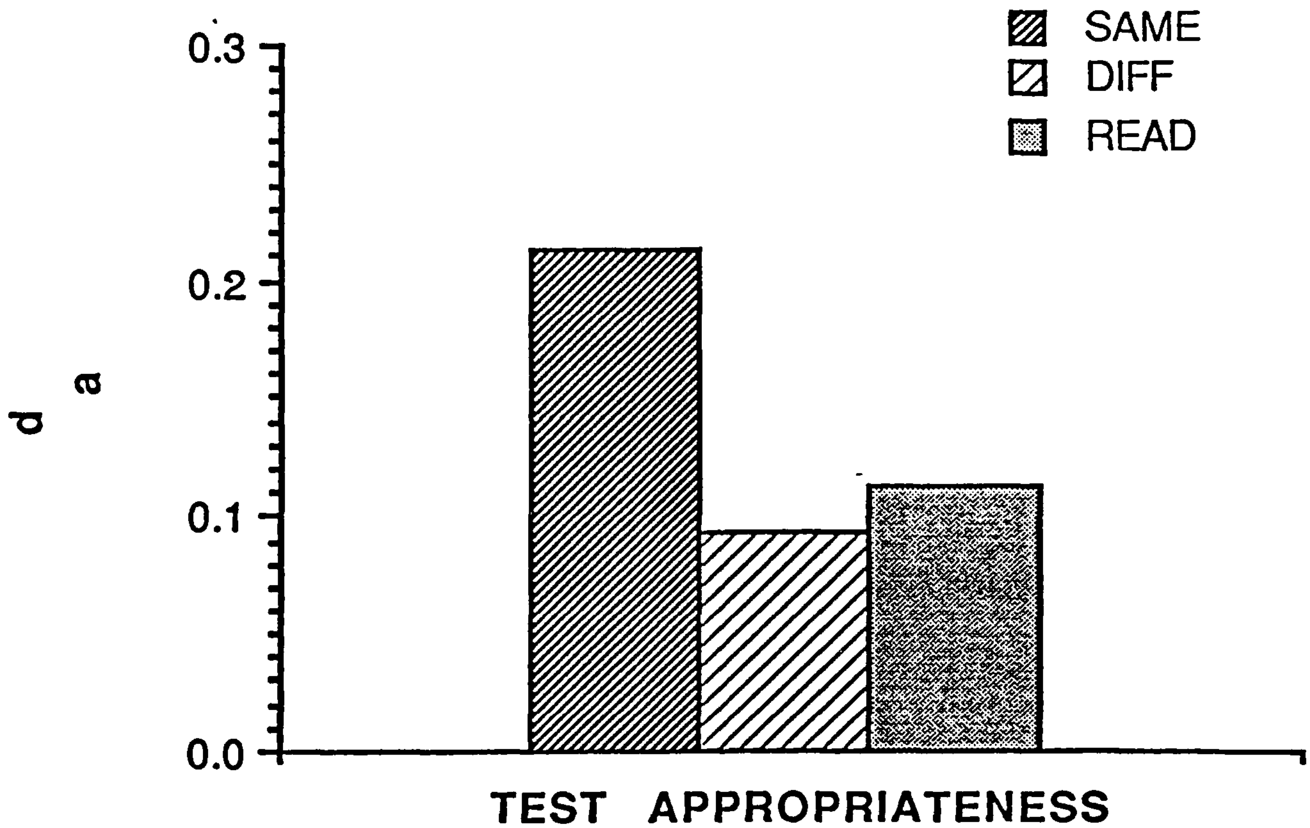

EXPERIMENT 2
RESPONSE LATENCY (sec.)
1.90
1.85
1.80
1.75
1.70
CORRECT
INCORRECT
RECOGNITION DECISION
LIST TYPE
OLD5
NEW

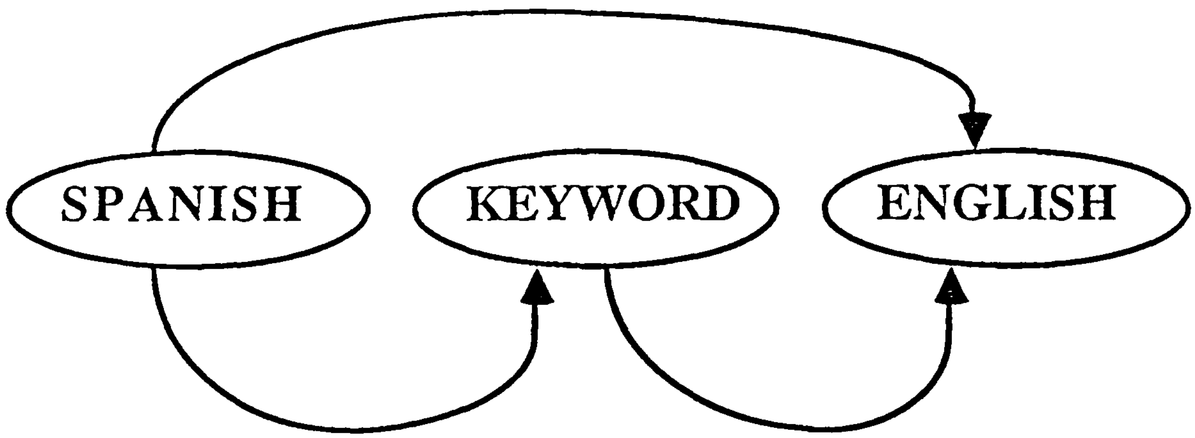
CALCULATOR				TELEPHONE		
7	8	9		1	2	3
4	5	6		4	5	6
1	2	3		7	8	9


RESPONSE LATENCY (sec.)
1.40
1.35
1.30
1.25
LIST TYPE
OLD DIGIT
OLD MOTOR
NEW

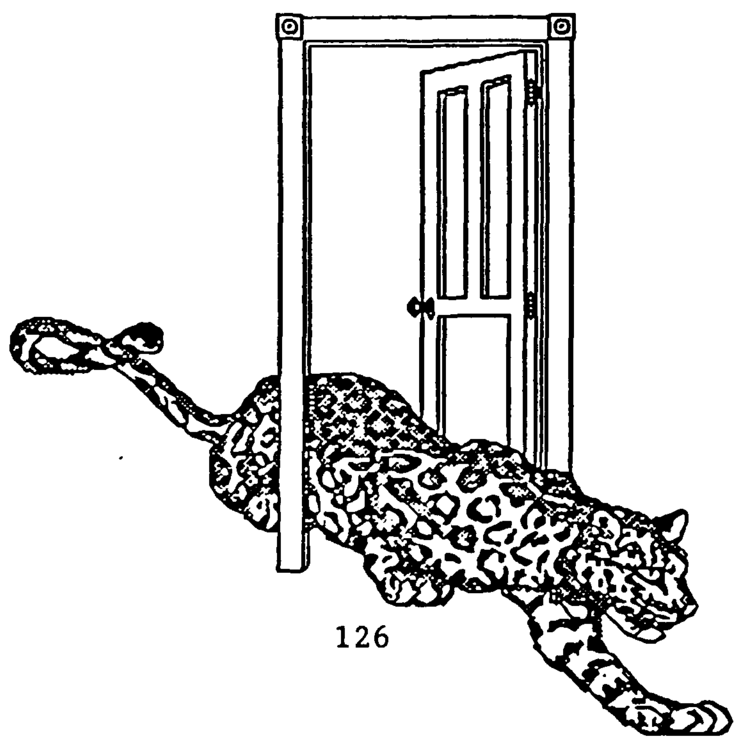

KEYPAD
ROW
READ
NEW
3.1
3.0
2.9
RESPONSE LATENCY (sec.)
TYPE OF ITEMS

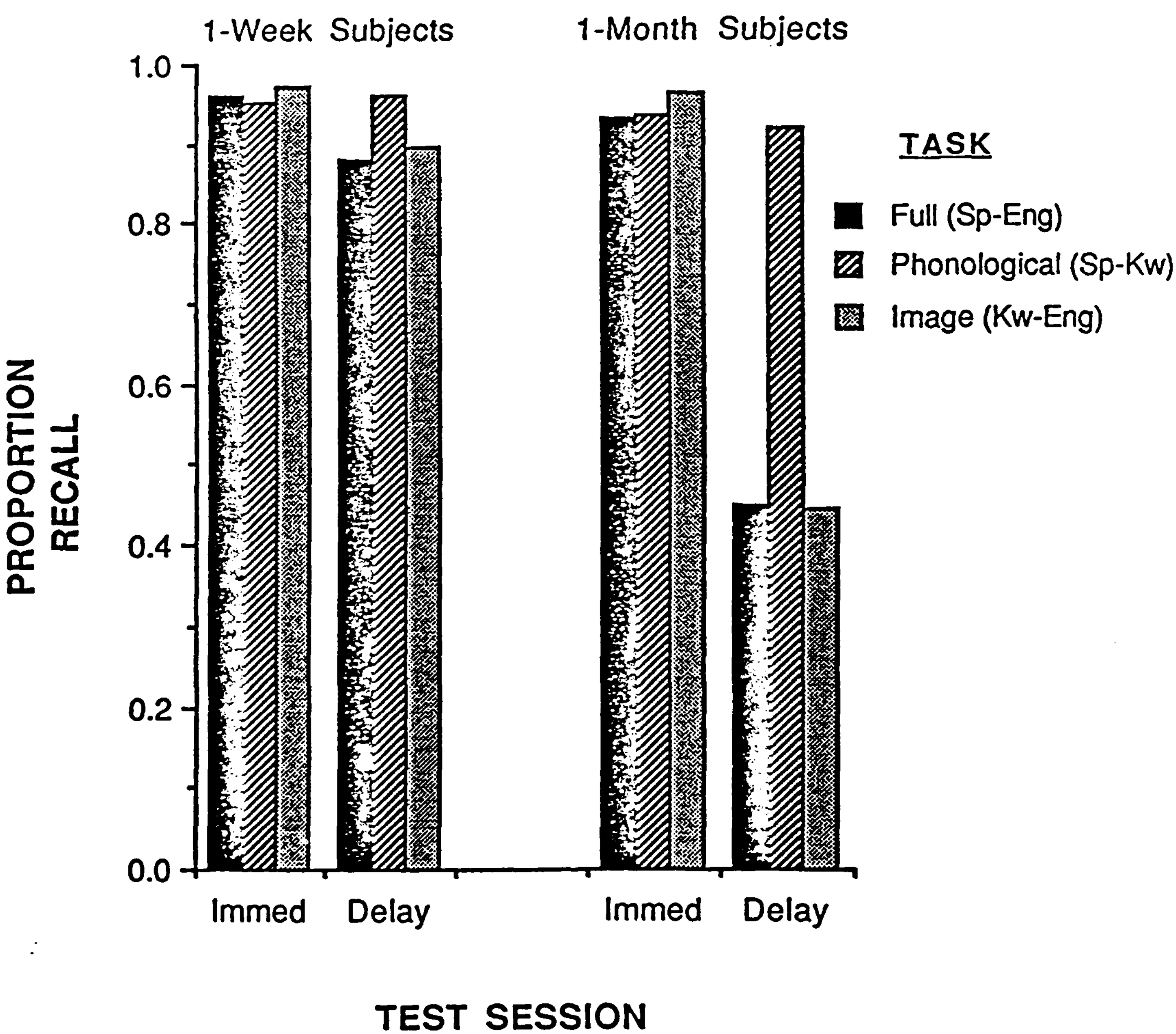

SAME
DIFF
READ
3.10
3.05
3.00
2.95
2.90
RESPONSE LATENCY (sec.)
TEST APPROPRIATENESS

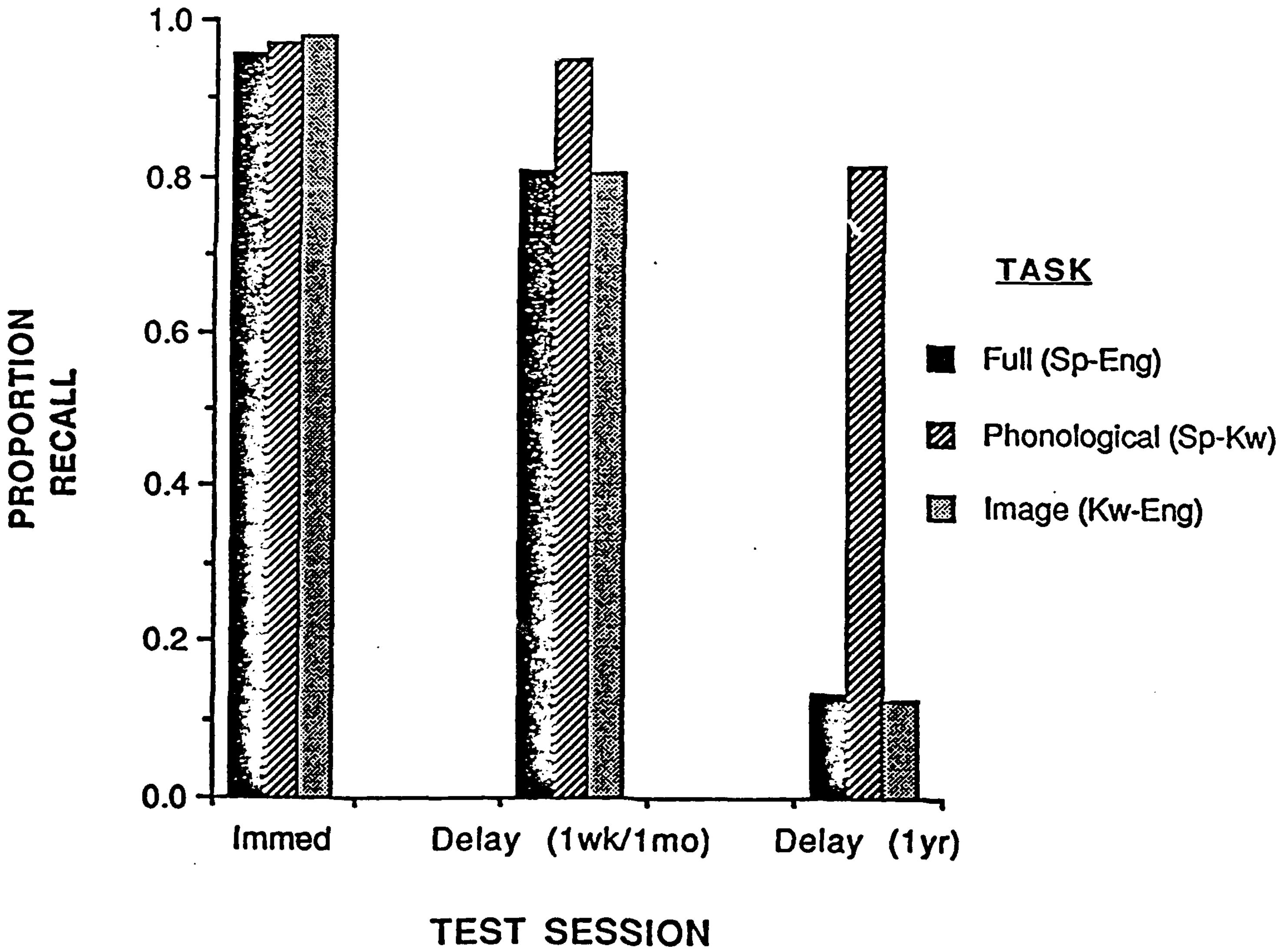

SAME
DIFF
READ
0.3
0.2
0.1
0.0
d
a
TEST APPROPRIATENESS

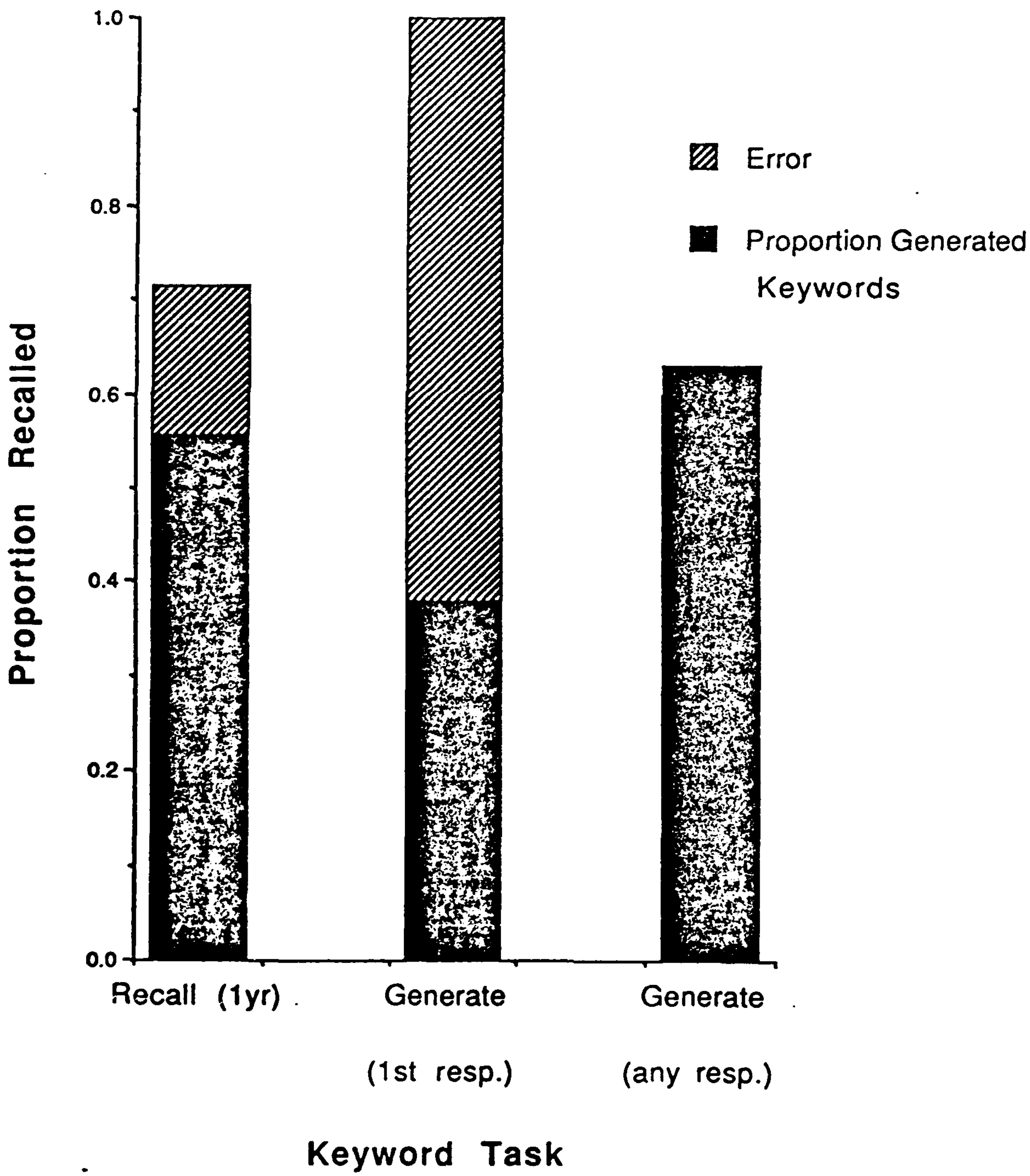


DORONICO **DOOR** **LEOPARD**

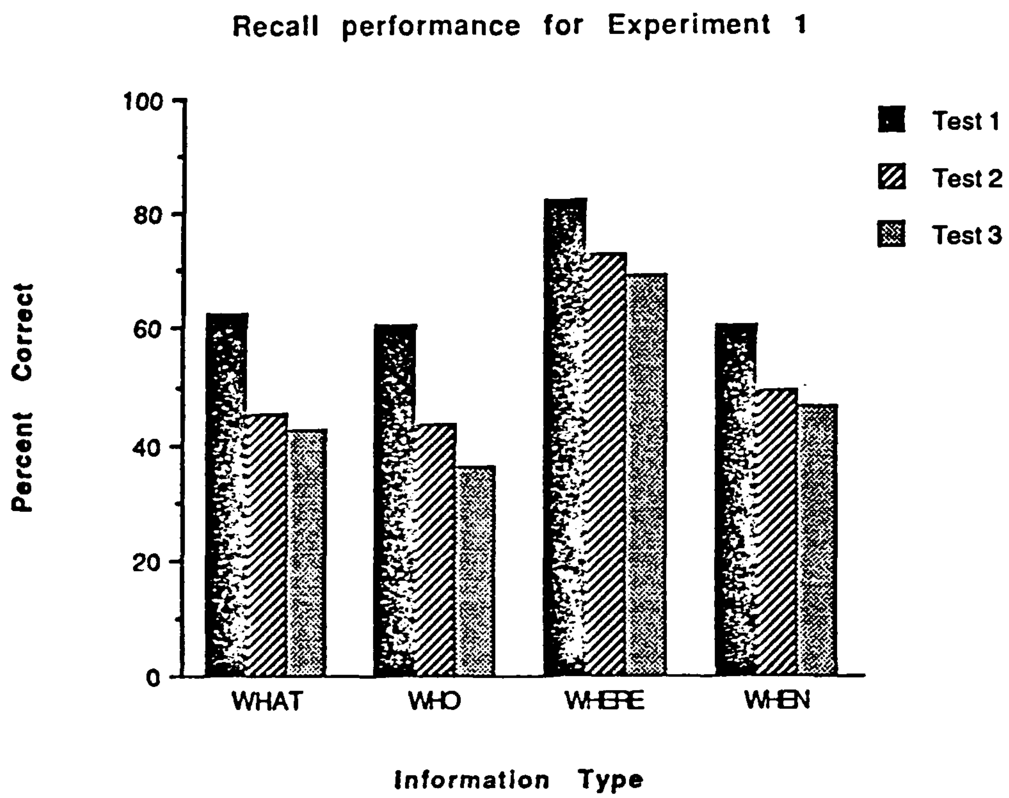



1-Week Subjects
1-Month Subjects
1.0
0.8
0.6
0.4
0.2
0.0
PROPORTION RECALL
TASK
Full (Sp-Eng)
Phonological (Sp-Kw)
Image (Kw-Eng)
Immed
Delay
Immed
Delay
TEST SESSION

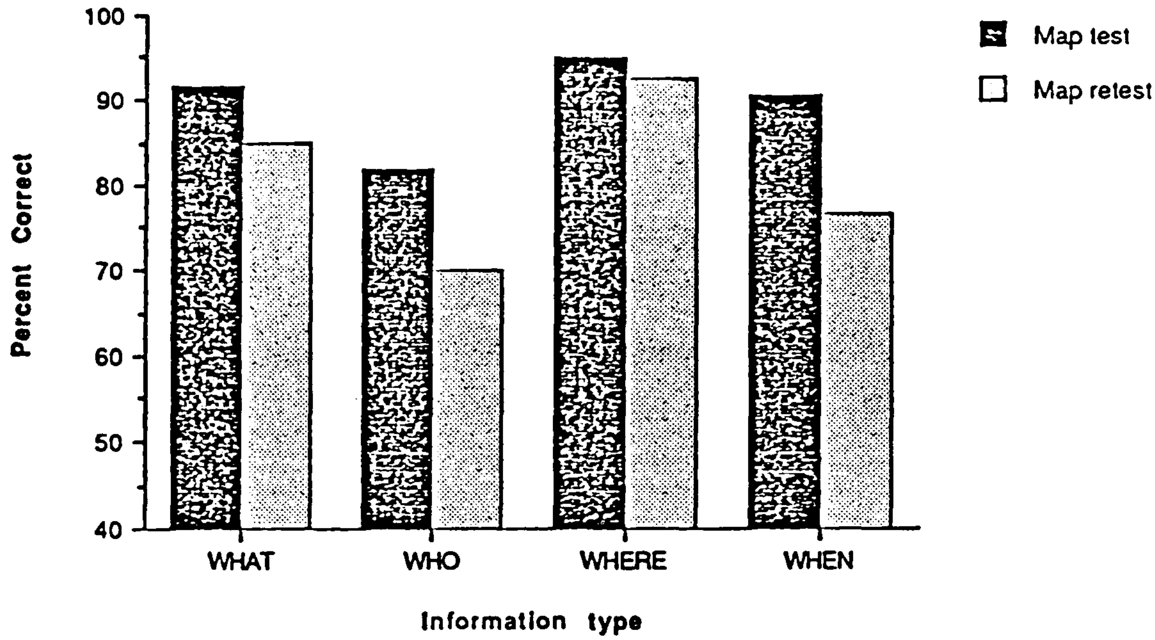

1.0
0.8
0.6
0.4
0.2
0.0
PROPORTION RECALL
Immed
Delay (1wk/1mo)
Delay (1yr)
TEST SESSION
TASK
Full (Sp-Eng)
Phonological (Sp-Kw)
Image (Kw-Eng)



1.0
0.8
0.6
0.4
0.2
0.0
Proportion Recalled
Error
Proportion Generated Keywords
Recall (1yr)
Generate (1st resp.)
Generate (any resp.)
Keyword Task


Recall performance for Experiment 1
100
80
60
40
20
0
Percent Correct
Test 1
Test 2
Test 3
WHAT
WHO
WHERE
WHEN
Information Type

Map test and retest results - Exp 2



Class listing test and retest results - Exp 2